

# SIDON BASIS

JAVIER CILLERUELO

**ABSTRACT.** Erdős conjectured the existence of an infinite Sidon sequence of positive integers which is also an asymptotic basis of order 3. We make progress towards this conjecture in several directions. First we prove the conjecture for all cyclic groups  $\mathbb{Z}_N$  with  $N$  large enough.

In second place we prove by probabilistic methods that there is an infinite  $B_2[2]$  sequence which is an asymptotic basis of order 3.

Finally we prove that for all  $\varepsilon > 0$  there is a Sidon sequence which is an asymptotic basis of order  $3 + \varepsilon$ , that is to say, any positive sufficiently large integer  $n$  can be written as a sum of 4 elements of the sequence, one of them smaller than  $n^\varepsilon$ .

## 1. INTRODUCTION

A sequence of positive integers  $A$  is a Sidon basis of order  $h$  if all the sums  $a + a'$ ,  $a \leq a'$ ,  $a, a' \in A$  are distinct (Sidon property) and if any positive integer  $n$ , large enough, can be written as a sum of  $h$  elements of  $A$  (asymptotic basis of order  $h$ ). It is not difficult to prove that there cannot be a Sidon basis of order 2, however Erdős stated the following ([8], [11], [12]):

**Conjecture 1.1.** *There is a Sidon basis of order 3.*

While we are not able to prove this we prove some results approaching it. The first one is the modular version of Conjecture 1.1.

**Theorem 1.1.** *For all  $N$  large enough, the cyclic group  $\mathbb{Z}_N$  contains a Sidon set  $S \subset \mathbb{Z}_N$  which is a basis of order 3 in  $\mathbb{Z}_N$ .*

We will prove this theorem in section §2. An important ingredient in the proof of Theorem 1.1 is a result of Granville, Shparlinski and Zaharescu [15] about distributions in the  $s$ -dimensional torus of points coming from curves in  $\mathbb{F}_p^r$ . We start section §2 with a weaker theorem that has a cleaner proof and is used in the proof of Theorem 1.2.

Theorem 1.2 is concerned with  $B_2[g]$  sequences, a natural generalization of Sidon sequences.

---

*Date:* April 25, 2013.

**Definition 1.** A sequence of positive integers  $A$  is a  $B_2[g]$  sequence if any integer  $n$  has at most  $g$  representations of the form  $n = a + a'$ ,  $a \leq a'$ ,  $a, a' \in A$ . The  $B_2[1]$  sequences are just the Sidon sequences.

Erdős claimed in [8] that there exists a  $B_2[g]$  sequence of positive integers which is an asymptotic basis of order 3 for some  $g$  and he asked for the minimum possible  $g$  (see also [19]). While Conjecture 1.1 would imply that the minimum is  $g = 1$ , we prove that  $g \leq 2$ .

**Theorem 1.2.** *There exists a  $B_2[2]$  sequence of positive integers which is an asymptotic basis of order 3.*

We next introduce a new generalization of basis that appears in the statement of our strongest approximation to Conjecture 1.1.

**Definition 2.** For any  $\varepsilon$ ,  $0 < \varepsilon < 1$  we say that  $A$  is an asymptotic basis of order  $h + \varepsilon$  if any sufficiently large positive integer  $n$  can be written as a sum of  $h + 1$  elements of  $A$ , one of them smaller than  $n^\varepsilon$ :

$$n = a_1 + \cdots + a_{h+1}, \quad a_1, \dots, a_{h+1} \in A, \quad a_{h+1} \leq n^\varepsilon.$$

We say that  $A$  is a Sidon basis of order  $h + \varepsilon$  if in addition it is a Sidon sequence.

**Theorem 1.3.** *For any  $\varepsilon > 0$  there exists a Sidon basis of order  $3 + \varepsilon$ . In other words: for any  $\varepsilon > 0$  there exists a Sidon sequence  $A$  of positive integers such that all positive integer  $n$ , large enough, can be written as*

$$n = a_1 + a_2 + a_3 + a_4, \quad a_1, a_2, a_3, a_4 \in A, \quad a_4 \leq n^\varepsilon.$$

We mention a couple of previous related results. Deshouillers and Plagne [6] constructed a Sidon basis of order 7 and Kiss [20] proved the existence of a Sidon basis of order to 5. See also note 1.1.

To prove Theorems 1.2 and 1.3 we will use the probabilistic method invented by Erdős and Renyi [10]. In the study of sequences satisfying certain additive properties they considered a probabilistic space  $\mathcal{S}(\gamma)$  of sequences of positive integers where all the events  $x \in A$  are independent and  $\mathbb{P}(x \in A) = x^{-\gamma}$ .

On the one hand an easy application of this method shows that if  $\gamma > 3/4$  then almost all sequences in  $\mathcal{S}(\gamma)$  are Sidon sequences (after we remove a finite number of elements from the sequence).

On the other hand Erdős and Tetali [13] proved that if  $\gamma < 1 - 1/h$ , then almost all sequences are asymptotic basis of order  $h$ . Thus, for any  $\gamma$  in the interval  $(3/4, 4/5)$  we have that almost all sequences in  $\mathcal{S}(\gamma)$  are simultaneously Sidon sequences and asymptotic basis of order 5. This is the argument used in [20].

In order to get basis of order  $3 + \varepsilon$  we need to take  $\gamma$  close to  $2/3$ . In this case the sequences are far from being Sidon sequences since we expect infinitely many repeated sums. A way to circumvent this obstacle is to remove the elements involved in such repetitions to obtain a true Sidon sequence. This general idea, which is called alteration method or deletion technique is standard in the probabilistic method (see [2]) and has been used previously in a similar context ([3, 5, 24]).

The main difficulty that appears when we apply the alteration method in our problem is that we have to prevent the destruction of all the representations of infinitely many integers  $n$ . So we need to prove that the number of removed elements involved in each representation is not too large.

As far as Theorem 1.2 is concerned, the standard application of the probabilistic method also proves that if  $\gamma > \frac{g+2}{2g+2}$  then with probability 1 a sequence in  $\mathcal{S}(\gamma)$  is a  $B_2[g]$  sequence (after we remove a finite number of elements). Thus, if  $\frac{5}{8} < \gamma < \frac{2}{3}$ , a random sequence in  $\mathcal{S}(\gamma)$  is, with probability 1, simultaneously a  $B_2[3]$  sequence and an asymptotic basis of order three. This result appears in [2] §8.6. To get a  $B_2[2]$  basis of order 3 we need to use a more involved argument.

In section §3 we explain in more detail the strategy of the proofs and the new ingredients we introduce: the vectorial sunflowers and the use of a modular Sidon basis.

The Sunflower Lemma was discovered by Erdős and Rado [9] and has many applications. In the probabilistic method it has been used to deal with dependent events when each event can be identified with a set. In our proofs it is more convenient to identify each event with a vector and then we have to use a vectorial version of the Sunflower Lemma. We refer to [1] for a recent study of other variants of sunflowers.

The modular Sidon basis are used as a trick to simplify the casuistry in the computations of the random variables appearing in the proofs. See section §3 for a more detailed explanation.

The proofs of Theorems 1.2 and 1.3 are quite similar, except that the last one is technically a little more involved. We prove them in sections §4 and §5 respectively, however we have left to the last section all the boring calculations of the expected values of the random variables appearing in the proofs.

**1.1. Note added on April 23, 2013.** Kiss, Rozgonyi and Sandor [21] have proved the existence of a Sidon sequence which is an asymptotic basis of order 4. Their result, which appeared in Arxiv only one day after the first version [4] of our preprint, has been obtained independently from our work. They use the method of Kim and Vu [22] to control the concentration of sums of dependent variables.

**1.2. General Notation.** Through the paper we will use the following notation:

- $f(n) \gg g(n)$  means that there exists  $C > 0$  such that  $f(n) > Cg(n)$  for  $n$  large enough. We observe that this includes the possibility that  $f(n) = 0$  for a finite number of positive integers  $n$ .

- $f(n) \ll g(n)$  means that there exists  $C > 0$  such that  $f(n) < Cg(n)$  for all  $n$ .
- $f(n) = o(g(n))$  means that  $f(n)/g(n) \rightarrow 0$  as  $n \rightarrow \infty$ .
- We write  $o_m(1)$  to mean a quantity tending to 0 as  $m \rightarrow \infty$ .

## 2. THE MODULAR VERSION OF THE CONJECTURE.

Our first theorem about modular Sidon basis is essentially contained in Theorem 1.1. However it includes the extra condition that  $s_1, s_2, s_3$  are pairwise distinct, which is convenient to be used in the proof of Theorem 1.2. Furthermore the proof is short and has an amusing relation with elliptic curves.

**Theorem 2.1.** *There exist infinitely many cyclic groups  $\mathbb{Z}_N$  containing a Sidon set  $S \subset \mathbb{Z}_N$  and such that any element  $x \in \mathbb{Z}_N$  can be written in the form*

$$(2.1) \quad x = s_1 + s_2 + s_3, \quad s_i \in S$$

with  $s_1, s_2, s_3$  pairwise distinct.

*Proof.* Ruzsa [23] observed that for all prime  $p$  and  $g$  a generator of  $\mathbb{F}_p^*$ , the set

$$S = \{(x, g^x) : x = 0, \dots, p-2\}$$

is a Sidon set in  $\mathbb{Z}_{p-1} \times \mathbb{Z}_p$ . Since  $\mathbb{Z}_{p-1} \times \mathbb{Z}_p$  is isomorphic to  $\mathbb{Z}_{(p-1)p}$  the set  $S$  provides an easy construction of a dense Sidon set in a cyclic group. We will prove that  $S$  is also a basis of order 3. In other words, that any element  $(a, b) \in \mathbb{Z}_{p-1} \times \mathbb{Z}_p$  can be written as

$$(2.2) \quad (a, b) = (x_1, g^{x_1}) + (x_2, g^{x_2}) + (x_3, g^{x_3}).$$

Indeed we will prove that the number of solutions of (2.2) is exactly the number of points  $(U, V)$ ,  $V \neq 0$  of the elliptic curve  $U^2 = 4V^3 + (bV + g^a)^2$  in  $\mathbb{F}_p$ .

We count, for any  $(a, b) \in \mathbb{Z}_{p-1} \times \mathbb{Z}_p$ , the number of solutions  $(x_1, x_2, x_3)$  of the system

$$(2.3) \quad x_1 + x_2 + x_3 \equiv a \pmod{p-1}$$

$$(2.4) \quad g^{x_1} + g^{x_2} + g^{x_3} \equiv b \pmod{p}.$$

This can be written as

$$g^{x_1} + g^{x_2} + g^{a-x_1-x_2} \equiv b \pmod{p},$$

which is also equivalent to

$$(2.5) \quad X + Y + \frac{\lambda}{XY} \equiv b \pmod{p}$$

after the change  $g^{x_1} = X$ ,  $g^{x_2} = Y$ ,  $g^a = \lambda$ . Now we make another change of variables:

$$X = \frac{2V^2}{U - bV - \lambda}, \quad Y = -\frac{\lambda}{V}.$$

Since  $XY \neq 0$  we have to add the condition  $V \neq 0$ ,  $U \neq bV + \lambda$ . With these restrictions the change of variables is bijective. Applying this lucky change of variables in (2.5) we have

$$\begin{aligned} \frac{2V^2}{U - bV - \lambda} - \frac{\lambda}{V} - \frac{U - bV - \lambda}{2V} &\equiv b \pmod{p} \\ \iff \frac{2V^2}{U - bV - \lambda} &\equiv \frac{U + bV + \lambda}{2V} \pmod{p} \\ \iff 4V^3 + (bV + \lambda)^2 &\equiv U^2 \pmod{p}. \end{aligned}$$

Each point of this elliptic curve (except  $(U, V) = (\pm\lambda, 0)$ ) corresponds to a solution  $(X, Y)$  of (2.5). By Hasse's Theorem [16] we know that the elliptic curve has  $p + O(\sqrt{p})$  points  $(U, V)$ .

We have to remove the solutions  $(x_1, x_2, x_3)$  of (2.3) such that  $x_i = x_j$  for some  $i \neq j$ . Suppose that  $x_1 = x_2$ . In that case the equation (2.5) is  $2X + \frac{\lambda}{X^2} \equiv b \pmod{p}$ , which is a cubic equation having at most three solutions. Thus, the number of solutions  $(x_1, x_2, x_3)$  of (2.3) with some repeated coordinates is at most 9 and the number of representations of  $(a, b)$  as a sum of three pairwise distinct elements of  $S$  is  $p + O(\sqrt{p})$ .  $\square$

Corollary 2.1 is a byproduct of the proof above and will be used in the proof of Theorem 1.3.

**Corollary 2.1.** *There exist infinitely many cyclic groups  $\mathbb{Z}_N$  containing a Sidon set  $S \subset \mathbb{Z}_N$  and such that any element  $x \in \mathbb{Z}_N$  can be written in the form*

$$x = s_1 + s_2 + s_3 + s_4, \quad s_i \in S$$

*with  $s_1, s_2, s_3, s_4$  pairwise distinct.*

*Proof.* We will see that the set  $S \subset \mathbb{Z}_{p-1} \times \mathbb{Z}_p \simeq \mathbb{Z}_{(p-1)p}$  described in Theorem 2.1 satisfies the conditions of Corollary 2.1. From the proof of Theorem 2.1 we know that the number of representations of  $(a, b)$  as

$$(2.6) \quad (a, b) = (x_1, g^{x_1}) + (x_2, g^{x_2}) + (x_3, g^{x_3}) + (0, 1), \quad x_i \neq x_j, \quad 1 \leq i < j \leq 3$$

is  $p + O(\sqrt{p})$ . We observe that all these representations of  $(a, b)$  satisfy the conditions of Corollary 2.1 except those with  $x_i = 0$  for some  $i = 1, 2, 3$ . In these cases the equation (2.5) is a quadratic equation and the number of these special representations is at most 6 for each  $(a, b)$ .  $\square$

**2.1. Proof of Theorem 1.1.** We need the following weaker version of a result of Granville, Shparlinski and Zaharesku [15]:

**Theorem 2.2.** *Let  $\mathcal{C}$  be a curve of degree  $d$  in  $\mathbb{F}_p^r$  which is absolutely irreducible in  $\mathbb{A}^r(\overline{\mathbb{F}}_p)$ . Let  $h : \mathcal{C} \rightarrow \mathbb{A}^s(\overline{\mathbb{F}}_p)$  be a function  $h(X) = (h_1(X), \dots, h_s(X))$  where  $h_i(X)$ ,  $i = 1, \dots, s$  are polynomial functions.*

*Assume also that there exists  $L = L(p) \rightarrow \infty$  such that  $c_1 = \dots = c_s = 0$  whenever  $|c_i| \leq L$ ,  $i = 1, \dots, s$  and  $c_1 h_1(X) + \dots + c_s h_s(X)$  is constant along  $\mathcal{C}$ .*

*Under these conditions, the set*

$$S = \left\{ \left( \frac{h_1(X)}{p}, \dots, \frac{h_s(X)}{p} \right) : X \in \mathcal{C} \right\}$$

*is well distributed in  $\mathbb{T}^s$  when  $p \rightarrow \infty$ .*

Proposition 2.1 is a consequence of Theorem 2.2.

**Proposition 2.1.** *Let  $p \equiv 1 \pmod{3}$  be a prime. For any integers  $r_1, r_2$ , let  $\mathcal{C}_{r_1, r_2}$  be the curve in  $\mathbb{F}_p^2$  defined by*

$$(2.7) \quad x_1^2 + x_2^2 + (x_1 + x_2 - r_1)^2 \equiv r_2 \pmod{p}.$$

*The set*

$$(2.8) \quad S_{r_1, r_2} = \left\{ \left( \frac{(x_1)_p}{p}, \frac{(x_2)_p}{p}, \frac{(x_1^2)_p}{p}, \frac{(x_2^2)_p}{p} \right) : (x_1, x_2) \in \mathcal{C}_{r_1, r_2} \right\}$$

*is well distributed in  $[0, 1]^4$  when  $p \rightarrow \infty$ . In particular, given  $c > 0$ , any box  $B \subset [0, 1]^4$  of size  $|B| > c$  contains an element of  $S_{r_1, r_2}$  if  $p$  is large enough.*

*Proof.* The equation (2.7) can be written as

$$(2.9) \quad 3(2x_1 + x_2 - r_1)^2 + (3x_2 - r_1)^2 \equiv 6r_2 - 2r_1^2 \pmod{p}.$$

Thus, the curve is absolutely irreducible if  $6r_2 - 2r_1^2 \not\equiv 0 \pmod{p}$ .

In the first place we consider the case  $6r_2 - 2r_1^2 \not\equiv 0 \pmod{p}$ .

We will prove that  $L = \sqrt{p}/3$  works for the condition of Theorem 2.2. Suppose that

$$(2.10) \quad c_1 x_1 + c_2 x_2 + c_3 x_1^2 + c_4 x_2^2 = c_0$$

for all  $(x_1, x_2) \in \mathcal{C}_{r_1, r_2}$ . From (2.7) we have that

$$x_1^2 = x_1(r_1 - x_2) - x_2^2 + r_1 x_2 + \frac{r_2 - r_1^2}{2}.$$

Substituting  $x_1^2$  in (2.10) by this expression we have that

$$c_1 x_1 + c_2 x_2 + c_3 \left( x_1(r_1 - x_2) - x_2^2 + r_1 x_2 + \frac{r_2 - r_1^2}{2} \right) + c_4 x_2^2 = c_0.$$

which is equivalent to

$$(c_1 + c_3(r_1 - x_2))x_1 = (c_3 - c_4)x_2^2 - (c_3r_1 + c_2)x_2 + c_0 + \frac{c_3(r_1^2 - r_2)}{2}.$$

We write this in a short way as  $P(x_2)x_1 = Q(x_2)$  with

$$\begin{aligned} P(x_2) &= -c_3x_2 + c_1 + c_3r_1. \\ Q(x_2) &= (c_3 - c_4)x_2^2 - (c_3r_1 + c_2)x_2 + c_0 + \frac{c_3(r_1^2 - r_2)}{2}. \end{aligned}$$

Multiplying (2.10) by  $4c_3$  and completing squares we have

$$(2c_3x_1 + c_1)^2 + 4c_2c_3x_2 + 4c_3c_4x_2^2 = 4c_3c_0 + c_1^2.$$

Multiplying by  $P(x_2)^2$  and using that  $P(x_2)x_1 = Q(x_2)$  we have

$$(2c_3Q(x_2) + c_1P(x_2))^2 + P(x_2)^2(4c_2c_3x_2 + 4c_3c_4x_2^2 - 4c_3c_0 - c_1^2) = 0.$$

This equality must be satisfied for all  $x_2$  corresponding to a point  $(x_1, x_2) \in \mathcal{C}_{r_1, r_2}$ . Since the left hand of the equality above is a polynomial in  $x_2$  of degree less than or equal to 4, this can only be possible if it is the zero polynomial. It is easy to check that the coefficient of  $x_2^4$  in the polynomial is

$$4c_3^2(c_3 - c_4)^2 + 4c_3^3c_4 = 4c_3^2(c_3^2 + c_4^2 - c_3c_4).$$

If  $c_3 \neq 0$  we have that  $c_3^2 + c_4^2 - c_3c_4 \equiv 0 \pmod{p}$ . The inequality

$$|c_3^2 + c_4^2 - c_3c_4| \leq 3L^2 < p$$

implies that  $c_3^2 + c_4^2 - c_3c_4 = 0 \implies c_3 = c_4 = 0$ . Thus  $c_3 = 0$  in any case.

Since  $x_1$  and  $x_2$  play the same role, we can proceed in the same way to deduce that  $c_4 = 0$ . Now we have to consider the possibility that  $c_1x_1 + c_2x_2 = c_0$  for any  $(x_1, x_2) \in \mathcal{C}_{r_1, r_2}$ . But this means that all the solutions of the curve  $\mathcal{C}_{r_1, r_2}$  lie on that line, which is impossible unless  $c_0 = c_1 = c_2 = 0$ .

We have proved that the conditions of Theorem 2.2 are satisfied when  $6r_2 - 2r_1^2 \not\equiv 0 \pmod{p}$  and then the sets  $S_{r_1, r_2}$  are well distributed in this case.

Assume now that  $6r_2 - 2r_1^2 \equiv 0 \pmod{p}$ .

We observe that in this case the curve (2.9) is not absolutely irreducible. Let  $\omega$  be a solution of  $\omega^2 + \omega + 1 \equiv 0 \pmod{p}$ , which exists because  $p \equiv 1 \pmod{3}$ . It is easy to check that the points  $(x_1, x_2)$  of (2.9) are those satisfying either  $x_1 = \omega x_2 + (1 - \omega)r_1$  or  $x_1 = -\omega x_2 + (1 + \omega)r_1$  and the curve (2.9) is the union of the two lines:

$$\begin{aligned} \mathcal{C}_{r_1}^+ &= \{(x_1, x_2) : x_1 = \omega x_2 + (1 - \omega)r_1 : x_1, x_2 \in \mathbb{F}_p\} \\ \mathcal{C}_{r_1}^- &= \{(x_1, x_2) : x_1 = -\omega x_2 + (1 + \omega)r_1 : x_1, x_2 \in \mathbb{F}_p\}. \end{aligned}$$

We use again Theorem 2.2 to prove that the set (2.8) is well distributed when  $(x_1, x_2)$  belongs to both curves. We will do the work for the first one (for the second one the task is similar).

It is clear that  $\mathcal{C}_{r_1}^+$  is absolutely irreducible because it has degree 1. We will prove that  $L = \sqrt{p}/3$  satisfies the condition of Theorem 2.2.

Suppose that there exist constants  $c_0, c_1, c_2, c_3, c_4$  such that

$$c_1x_1 + c_2x_2 + c_3x_1^2 + c_4x_2^2 = c_0$$

for all  $(x_1, x_2) \in \mathcal{C}_{r_1}^+$ . In this case we would have

$$c_1(\omega x_2 + (1 - \omega)r_1) + c_2x_2 + c_3(\omega x_2 + (1 - \omega)r_1)^2 + c_4x_2^2 = c_0$$

for all  $x_2 \in \mathbb{F}_p$  which is not possible if the coefficient of  $x_2^2$  is not 0. So  $c_4 = -c_3\omega^2$  and using that  $\omega^2 + 1 = -\omega$  we have that  $c_4 - c_3 = -c_3\omega^2 - c_3 = c_3\omega$ . Thus

$$(c_4 - c_3)^2 + c_3(c_4 - c_3) + c_3^2 = (c_3\omega)^2 + c_3^2\omega + c_3^2 = c_3^2(\omega^2 + \omega + 1) \equiv 0 \pmod{p}.$$

On the one hand we know that

$$|(c_4 - c_3)^2 + c_3(c_4 - c_3) + c_3^2| \leq (2L)^2 + L(2L) + L^2 \leq 7L^2 < p.$$

This implies that  $(c_4 - c_3)^2 + c_3(c_4 - c_3) + c_3^2 = 0$  which implies that  $c_3 = c_4 - c_3 = 0$  and then  $c_3 = c_4 = 0$ .

Then the coefficient of  $x_2$  must be also 0, so  $c_2 = -c_1\omega$  and we have that

$$c_2^2 - c_1c_2 + c_1^2 = c_1^2\omega^2 + c_1^2\omega + c_1^2 = c_1^2(\omega^2 + \omega + 1) \equiv 0 \pmod{p}.$$

On the other hand

$$|c_2^2 - c_1c_2 + c_1^2| \leq 3L^2 < p.$$

This implies that  $c_2^2 - c_1c_2 + c_1^2 = 0$  which implies that  $c_1 = c_2 = 0$ . So  $c_0 = \dots = c_4$ .

We have proved that the conditions of Theorem 2.2 are also satisfied when  $6r_2 - 2r_1^2 \equiv 0 \pmod{p}$  for  $\mathcal{C}_{r_1}^+$  (and similarly for  $\mathcal{C}_{r_1}^-$ ) and then the set  $S_{r_1, r_2}$  is well distributed in all the cases.  $\square$

**2.2. End of the proof of Theorem 1.1.** Erdős and Turan [14] showed that the set

$$A = \{x + (x^2)_p(2p) : x = 0, \dots, p-1\}$$

is a Sidon set of integers for any odd prime  $p$ .

For given  $N$  let  $p$  be a prime such that  $p \equiv 1 \pmod{3}$  and  $4p^2 < N < 5p^2$ . This prime exists if  $N$  is large enough. Since  $A \subset [0, 2p^2) \subset [0, N/2)$ , the set  $A$  is a Sidon set in  $\mathbb{Z}_N$ . We will prove that  $A$  is also a basis of order 3 in  $\mathbb{Z}_N$ .



We observe that for any integer  $K$ , the set of integers of the form

$$(2.11) \quad r_1 + r_2(2p), \quad K \leq r_1, r_2 \leq \frac{5p-1}{2} + K$$

covers an interval of length  $5p^2$ . This is clear for  $K = 0$  and, by translation, for all  $K$ . Since  $5p^2 > N$ , in order to prove that  $A$  is a basis of order 3 in  $\mathbb{Z}_N$  it is enough to prove that any element of the form (2.11) can be written as a sum of 3 elements of  $A$ . We will take  $K = \lceil p/4 \rceil$  through the proof.

For each  $(r_1, r_2)$  we consider the box  $B_{r_1, r_2} \subset [0, 1]^4$  of all points  $(y_1, y_2, y_3, y_4)$  satisfying the restrictions

$$\left| y_1 - \frac{r_1}{3p} \right|, \left| y_2 - \frac{r_1}{3p} \right|, \left| y_3 - \frac{r_2}{3p} \right|, \left| y_4 - \frac{r_2}{3p} \right| \leq \frac{K}{12p}.$$

We have to check that  $0 < y_i < 1$ ,  $i = 1, \dots, 4$  and then that  $B_{r_1, r_2} \subset [0, 1]^4$ . Indeed, since  $p \geq 7$ , we have

$$y \leq \frac{r_i}{3p} + \frac{K}{12p} \leq \frac{\frac{5p-1}{2} + K}{3p} + \frac{K}{12p} < \frac{5p-1}{6p} + \frac{5K}{12p} \leq \frac{5p-1}{6p} + \frac{5(p+3)}{48p} \leq \frac{45p+7}{48p} < 1$$

$$\text{and } y \geq \frac{r_1}{3p} - \frac{K}{12p} \geq \frac{K}{3p} - \frac{K}{12p} > 0.$$

The size of this box is  $|B_{r_1, r_2}| \geq \left(\frac{K}{12p}\right)^4 > 48^{-4}$  and then Proposition 2.1 implies that for  $p$  large enough there exists an element, say  $\left(\frac{x_1}{p}, \frac{x_2}{p}, \frac{(x_1^2)_p}{p}, \frac{(x_2^2)_p}{p}\right)$ , with  $0 \leq x_1, x_2 \leq p-1$ ,  $(x_1, x_2) \in \mathcal{C}_{r_1, r_2}$  satisfying

$$\left| \frac{x_1}{p} - \frac{r_1}{3p} \right|, \left| \frac{x_2}{p} - \frac{r_1}{3p} \right|, \left| \frac{(x_1^2)_p}{p} - \frac{r_2}{3p} \right|, \left| \frac{(x_2^2)_p}{p} - \frac{r_2}{3p} \right| \leq \frac{K}{12p}.$$

Since  $(x_1, x_2) \in \mathcal{C}_{r_1, r_2}$  there exists an integer  $x_3$ ,  $0 \leq x_3 \leq p-1$  satisfying

$$\begin{aligned} x_1 + x_2 + x_3 &\equiv r_1 \pmod{p} \\ x_1^2 + x_2^2 + x_3^2 &\equiv r_2 \pmod{p}. \end{aligned}$$

Let  $m$  be such that  $x_1 + x_2 + x_3 = r_1 + mp$ . We have

$$\begin{aligned} |m| &\leq \left| \frac{x_1}{p} - \frac{r_1}{3p} \right| + \left| \frac{x_2}{p} - \frac{r_1}{3p} \right| + \left| \frac{x_3}{p} - \frac{r_1}{3p} \right| \\ &\leq \frac{K}{12p} + \frac{K}{12p} + \max\left(\frac{r_1}{3p}, 1 - \frac{r_1}{3p}\right) \\ &\leq \frac{K}{6p} + \max\left(\frac{5p/2 + K}{3p}, 1 - \frac{K}{3p}\right) \\ &\leq \max\left(\frac{5p + 3K}{6p}, 1 - \frac{K}{6p}\right) < 1, \end{aligned}$$

since  $K = \lceil \frac{p}{4} \rceil$  and  $p \geq 7$ . This proves that indeed  $x_1 + x_2 + x_3 = r_1$ . The same argument proves that  $(x_1^2)_p + (x_2^2)_p + (x_3^2)_p = r_2$ . Thus we have

$$r_1 + r_2(2p) = x_1 + (x_1^2)_p(2p) + x_2 + (x_2^2)_p(2p) + x_3 + (x_3^2)_p(2p),$$

that is what we wanted to prove.

### 3. THE PROBABILISTIC METHOD WITH SOME NEW TOOLS

The proofs of Theorems 1.2 and 1.3 are based on the probabilistic method introduced by Erdős and Renyi [10] to study sequences satisfying certain arithmetic properties. The book [2] is the most complete reference on the probabilistic method and [17] is a classic reference for the probabilistic method applied to sequences of integers.

For a given  $\gamma$ , with  $0 < \gamma < 1$ , Erdős and Renyi introduced the probabilistic space  $\mathcal{S}(\gamma)$  of all sequences of positive integers  $A$  such that all the events  $x \in A$  are independent and  $\mathbb{P}(x \in A) = x^{-\gamma}$ .

Generally speaking the goal is to prove that a sequence  $A$  in  $\mathcal{S}(\gamma)$  satisfies certain arithmetic property (or properties) with high probability. To be more precise, we consider certain families  $\Omega_n$  of sets of positive integers and the random families

$$\Omega_n(A) = \{\omega \in \Omega_n : \omega \subset A\}$$

generated by a random sequence  $A$  in  $\mathcal{S}(\gamma)$ . Typically we are interested in the random variable

$$X_n(A) = |\Omega_n(A)| = \sum_{\omega \in \Omega_n} I(\omega \subset A).$$

For example if

$$(3.1) \quad \Omega_n = \{\omega = \{x_1, x_2, x_3\} : x_1 + x_2 + x_3 = n\},$$

the random variable  $X_n(A)$  counts the number of representation of  $n$  as a sum of three elements of a random sequence  $A$  in  $\mathcal{S}(\gamma)$ . In general we are interested in proving that  $X_n(A)$  satisfies a certain property  $P_n$ . The standard strategy is to first prove that

$$(3.2) \quad \sum_n \mathbb{P}(X_n(A) \text{ does not satisfies } P_n) < \infty$$

and then apply Borel-Cantelli lemma to deduce that with probability 1, the random variable  $X_n(A)$  satisfies property  $P_n$  for all  $n$  large enough.

We will modify the probabilistic space  $\mathcal{S}(\gamma)$  to force that all the elements of  $A$  lie in some residue classes  $s \in S \pmod{N}$  for some  $S \subset \mathbb{Z}_N$  satisfying suitable conditions. At the end of this section we explain the advantage of this modification. We will write  $x \equiv S \pmod{N}$  to mean that  $x \equiv s \pmod{N}$  for some  $s \in S$ .

Also it is technically more convenient to introduce a parameter  $m$  to force that the elements of  $A$  are greater than a fixed  $m$ . This idea was introduced before in [5] and allows us to bound (3.2) by a quantity which is  $o_m(1)$ . At a later step we take  $m$  as large as we want.

**Definition 3.** *Let  $S$  be a non empty set of a cyclic group  $\mathbb{Z}_N$ . For a given  $\gamma$ ,  $0 < \gamma < 1$  and a given positive integer  $m$ , let  $\mathcal{S}_m(\gamma; s \bmod N)$  be the probabilistic space of all sequences of positive integers  $A$  such that all the events  $x \in A$  are independent and such that*

$$\mathbb{P}(x \in A) = \begin{cases} x^{-\gamma} & \text{if } x \equiv S \pmod{N} \text{ and } x > m \\ 0 & \text{otherwise.} \end{cases}$$

Since  $X_n(A)$  is a sum of boolean variables we expect that  $X_n(A)$  is concentrated around its expected value,  $\mu_n = \mathbb{E}(X_n(A))$ , with high probability.

When the variables  $I(\omega \in A)$  are independent (the sets  $\omega \in \Omega_n$  are disjoint), Chernoff's theorem is enough to prove that  $X_n(A)$  is strongly concentrated around  $\mu_n$ . However, when the sets in  $\Omega_n$  are not disjoint, as in the example (3.1), the study of the concentration is more involved.

It is expected, however, that if the dependent events have small correlation we still have enough concentration. Janson's inequality [18] serves our purpose for the lower tail:

**Theorem 3.1** (Janson's inequality). *Let  $\Omega$  be a family of sets and let  $A$  be a random subset. Let  $X(A) = |\{\omega \in \Omega : \omega \subset A\}|$  with finite expected value  $\mu = \mathbb{E}(X(A))$ . Then*

$$\mathbb{P}(X \leq (1 - \varepsilon)\mu) \leq \exp(-\varepsilon^2 \mu^2 / (2\mu + \Delta(\Omega)))$$

where

$$\Delta(\Omega) = \sum_{\substack{\omega, \omega' \in \Omega \\ \omega \sim \omega'}} \mathbb{P}(\omega, \omega' \subset A)$$

and  $\omega \sim \omega'$  means that  $\omega \cap \omega' \neq \emptyset$  and  $\omega \neq \omega'$ . In particular, if  $\Delta(\Omega) < \mu$  we have that

$$\mathbb{P}(X \leq \mu/2) \leq \exp(-\mu/12).$$

To deal with the upper tail Erdős and Tetali [13] introduced the Sunflowers trick.

**3.1. Sunflowers and vectorial Sunflowers.** A collection of sets  $S_1, \dots, S_k$  forms a sunflower if there exists a set  $C$  such that  $S_i \cap S_j = C$  for any  $i \neq j$ . The sets  $S_i \setminus C$  are the petals and  $C$  is the core of the sunflower. Erdős and Rao [9] proved the following interesting lemma.

**Lemma 3.1** (Sunflower lemma). *Let  $\Omega$  a family of  $h$ -sets. If  $\Omega$  does not contain a sunflower of  $k$  petals then  $|\Omega| \leq h!(k-1)^h$ .*

We will work with a variant of the Sunflower lemma which deals with vectors instead of sets. The reason is that in our proofs sometimes it will be more convenient to work with families  $\Omega$  of vectors (instead of sets).

**Definition 4.** For a given vector  $\bar{x} = (x_1, \dots, x_h)$  we define  $\text{Set}(\bar{x}) = \{x_1, \dots, x_h\}$ . We say that a collection of  $k$  distinct vectors  $\bar{x}_j$ ,  $j = 1, \dots, k$  forms a disjoint set of  $k$  vectors ( $k$ -d.s.v. for short) if  $\text{Set}(\bar{x}_j) \cap \text{Set}(\bar{x}_{j'}) = \emptyset$  for any  $j \neq j'$ .

**Definition 5.** We say that  $k$  distinct vectors with  $h$  coordinates form a vectorial sunflower (of  $k$  petals) if for some  $I \subset [h]$  the following two conditions are satisfied:

- For all  $i \in I$  all the vectors have the same  $i$ -th coordinate.
- The set of vectors obtained by removing all the  $i$ -th coordinates,  $i \in I$ , forms a  $k$ -d.s.v.

We say that the vectorial sunflower is of type  $I$ .

We observe that a vectorial sunflower (of  $k$  petals) of type  $I = \emptyset$  is a  $k$ -d.s.v. The example below forms a vectorial sunflower (of 4 petals) of type  $I = \{2, 5\}$ .

$$\begin{aligned}\bar{x}_1 &= (7, \mathbf{7}, 1, 13, \mathbf{8}) \\ \bar{x}_2 &= (17, \mathbf{7}, 6, 6, \mathbf{8}) \\ \bar{x}_3 &= (8, \mathbf{7}, 18, 8, \mathbf{8}) \\ \bar{x}_4 &= (11, \mathbf{7}, 4, 5, \mathbf{8})\end{aligned}$$

We need a vectorial version of Lemma 3.1.

**Lemma 3.2** (Vectorial sunflower lemma). *Let  $\Omega$  be a family of vectors of  $h$  coordinates. If  $\Omega$  does not contain a vectorial sunflower of  $k$  petals then  $|\Omega| \leq h!((h^2 - h + 1)k)^h$ .*

*Proof.* Suppose that  $|\Omega| > h!((h^2 - h + 1)(k - 1))^h$ . For any  $\bar{x} = (x_1, \dots, x_h) \in \Omega$  we consider the set  $\text{Set}_h(\bar{x}) = \{hx_1 + 1, hx_2 + 2, \dots, hx_h + h\}$  and the family  $\hat{\Omega} = \{\text{Set}_h(\bar{x}) : \bar{x} \in \Omega\}$ . The sunflower lemma of Erdős-Rao applied to  $\hat{\Omega}$  implies that there exists a classical sunflower with  $(h^2 - h + 1)(k - 1) + 1$  petals, say  $\text{Set}_h(\bar{x}_1), \dots, \text{Set}_h(\bar{x}_{(h^2 - h + 1)(k - 1) + 1})$ . It is clear that from these sets we can recover the corresponding vectors  $\bar{x}_1, \dots, \bar{x}_{(h^2 - h + 1)(k - 1) + 1}$  which satisfy the following conditions:

- There exists  $I \subset \{1, \dots, h\}$  such that for each  $i \in I$  all the  $(h^2 - h + 1)(k - 1) + 1$  vectors have the same  $i$ -th coordinate.
- For each  $i \notin I$ , the  $i$ -th coordinates of all these vectors are pairwise distinct.

We observe that the conditions above are not enough to make sure that the vectors form a vectorial sunflower. We will proof, however, that the set  $\{\bar{x}_1, \dots, \bar{x}_{(h^2 - h + 1)(k - 1) + 1}\}$  contains a vectorial sunflower of  $k$  petals.

Select one vector, say  $\bar{x}_1$ . We know that if  $i \notin I$ , the  $i$ -th coordinate of  $\bar{x}_1$  cannot be equal to the  $i$ -th coordinate of a distinct vector. However it may be equal to a different  $i'$ -th coordinate ( $i' \notin I$ ) of a distinct vector. We observe that for each  $i \notin I$  and for each  $i' \notin I$ ,  $i' \neq i$  there is at most one such vector.

We remove, for each  $i \notin I$  and for each  $i' \notin I$ ,  $i' \neq i$ , that vector (if it exists). Thus removing at most  $h(h-1)$  vectors we make sure that for all  $i \notin I$ , the  $i$ -th coordinate of  $x_1$  is not equal to any  $i'$ -th coordinate ( $i' \notin I$ ,  $i' \neq i$ ) of a distinct vector.

Now we select a second vector and proceed as above. Since the number of original vectors was  $(h(h-1)+1)(k-1)+1$  we can select at least  $k$  vectors in this way forming a vectorial sunflower of  $k$  petals.  $\square$

Typically we will deal with families of vectors  $\Omega$  and with the corresponding random families  $\Omega(A) = \{\bar{x} \in \Omega : \text{Set}(\bar{x}) \subset A\}$ .

**Corollary 3.1.** *Let  $\Omega_n$  be a sequence of families of vectors of  $h$  coordinates. Suppose that with probability  $1 - o_m(1)$  the random families  $\Omega_n(A)$  do not contain vectorial sunflowers of  $K$  petals for any  $n$ . Then, with probability  $1 - o_m(1)$  it holds that  $|\Omega_n(A)| \leq h!((h^2 - h + 1)K)^h$  for all  $n$ .*

The following proposition will be used several times in the proofs of Theorems 1.2 and 1.3.

**Proposition 3.1.** *Let  $\{\Omega_n\}$  be a sequence of families of vectors and  $\{\Omega_n(A)\}$  the corresponding random family where  $A$  is a random sequence in  $\mathcal{S}_m(\gamma, s \pmod{N})$ . Suppose that there is  $\delta > 0$  such that  $\mathbb{E}(|\Omega_n(A)|) \ll (n+m)^{-\delta}$ . If  $K > 1/\delta$  then*

$$\mathbb{P}(\Omega_n(A) \text{ contains a } K\text{-d.s.v. for some } n) = o_m(1).$$

*Proof.*

$$\begin{aligned} \mathbb{P}(\Omega_n(A) \text{ contains a } K\text{-d.s.v.}) &\leq \sum_{\substack{\bar{x}_1, \dots, \bar{x}_K \in \Omega_n \\ \text{form a } K\text{-d.s.v.}}} \mathbb{P}(\text{Set}(\bar{x}_1), \dots, \text{Set}(\bar{x}_K) \subset A) \\ &= \sum_{\substack{\bar{x}_1, \dots, \bar{x}_K \in \Omega_n \\ \text{form a } K\text{-d.s.v.}}} \mathbb{P}(\text{Set}(\bar{x}_1) \subset A) \cdots \mathbb{P}(\text{Set}(\bar{x}_K) \subset A) \\ &\leq \frac{1}{K!} \left( \sum_{\bar{x} \in \Omega_n} \mathbb{P}(\text{Set}(\bar{x}) \subset A) \right)^K \\ &= \frac{\mathbb{E}(|\Omega_n(A)|)^K}{K!} \ll \frac{(n+m)^{-\delta K}}{K!}. \end{aligned}$$

Then,

$$\begin{aligned} \mathbb{P}(\Omega_n(A) \text{ contains a K-d.s.v. for some } n) &\ll \sum_n \mathbb{P}(\Omega_n(A) \text{ contains a K-d.s.v.}) \\ &\ll \sum_n \frac{(n+m)^{-\delta K}}{K!} = o_m(1). \end{aligned}$$

□

**3.2. The modular trick.** Before we explain the strategy of the proof of Theorem 1.2 we advance that we will deal with sums of the form

$$(3.3) \quad \sum_{\substack{\bar{x}=(x_1,\dots,x_8) \\ x_1+x_2+x_3=n \\ x_1+x_4=x_5+x_6=x_7+x_8 \\ \{x_1,x_4\} \neq \{x_5,x_6\} \neq \{x_7,x_8\}}} \mathbb{P}(x_1, \dots, x_8 \in A)$$

where  $A$  is a random sequence. If the coordinates of a vector  $\bar{x}$  are pairwise distinct, then  $\mathbb{P}(x_1, \dots, x_8 \in A) = \prod_{i=1}^8 \mathbb{P}(x_i \in A)$  and the computation of (3.3) is straightforward. Unfortunately we have also to consider those vectors with repeated coordinates. There are many patterns to consider and the computation of the sum above would be hard in a standard probabilistic space  $\mathcal{S}(\gamma)$ . To reduce this unpleasant task we will restrict the sequences  $A$  to be in some residue classes  $s \in S, \pmod{N}$  for some  $S \subset \mathbb{Z}_N$  given in Theorem 2.1. This trick will simplify a lot the casuistry of the possible coincidences between the coordinates in the proofs of Lemmas 6.5 and 6.8.

#### 4. $B_2[2]$ SEQUENCES WHICH ARE ASYMPTOTIC BASIS OF ORDER 3

In this section we prove Theorem 1.2.

**4.1. Strategy of the proof.** We start by fixing a cyclic group  $\mathbb{Z}_N$  and a set  $S \subset \mathbb{Z}_N$  satisfying the conditions of Theorem 2.1. Throughout this section we will consider the probabilistic space  $\mathcal{S}_m(7/11; S \pmod{N})$ . Indeed, it would work any  $\gamma$ ,  $\frac{5}{8} < \gamma < \frac{2}{3}$ . We consider the sequence of sets

$$Q_n = \left\{ \omega = \{x_1, x_2, x_3\} : x_1 + x_2 + x_3 = n, x_i \not\equiv x_j \pmod{N}, i \neq j \right\}.$$

Given a sequence of positive integers  $A$  we define, for each  $n$ , the set

$$Q_n(A) = \{\omega \in Q_n : \omega \subset A\}.$$

**Definition 6** (Lifting process). *The  $B_2[2]$ -lifting process of a sequence  $A$  consists in removing from  $A$  those elements  $a_1 \in A$  such that there exist  $a_2, a_3, a_4, a_5, a_6 \in A$  with  $a_1 + a_2 = a_3 + a_4 = a_5 + a_6$  and  $\{a_1, a_2\} \neq \{a_3, a_4\} \neq \{a_5, a_6\}$ .*

We denote by  $A_{B_2[2]}$  the surviving elements of  $A$  after this process. The sequence  $A_{B_2[2]}$  clearly is a  $B_2[2]$  sequence.

We define

$$\begin{aligned} T_n &= \{ \bar{x} = (x_1, \dots, x_8) : \bar{x} \text{ satisfies } \text{cond}(T_n) \} \quad \text{where} \\ \text{cond}(T_n) &:= \begin{cases} \{x_1, x_2, x_3\} \in Q_n \\ x_1 + x_4 = x_5 + x_6 = x_7 + x_8, & \{x_1, x_4\} \neq \{x_5, x_6\} \neq \{x_7, x_8\} \\ x_1 \equiv x_5 \equiv x_7 \pmod{N}, & x_4 \equiv x_6 \equiv x_8 \pmod{N}. \end{cases} \end{aligned}$$

We define also

$$T_n(A) = \{ \bar{x} \in T_n : \text{Set}(\bar{x}) \subset A \}.$$

We will see that  $|T_n(A)|$  is an upper bound for the number of the representations of  $n$  counted in  $Q_n(A)$  that are destroyed in the  $B_2[2]$ -lifting process of  $A$  defined above.

Suppose that  $\omega = \{x_1, x_2, x_3\} \in Q_n(A)$  contains an element, say  $x_1$ , which is removed in the  $B_2[2]$ -lifting process. Then there exist  $x_4, x_5, x_6, x_7, x_8 \in A$  such that  $x_1 + x_4 = x_5 + x_6 = x_7 + x_8$  with  $\{x_1, x_4\} \neq \{x_5, x_6\} \neq \{x_7, x_8\}$ . On the other hand, since all  $x_i \equiv S \pmod{N}$  and  $S$  is a Sidon set in  $\mathbb{Z}_N$ , interchanging  $x_5$  with  $x_6$  and  $x_7$  with  $x_8$  if needed, we have that  $x_1 \equiv x_5 \equiv x_7 \pmod{N}$  and  $x_4 \equiv x_6 \equiv x_8 \pmod{N}$ . Thus, any  $\omega \in Q_n(A)$  destroyed in the  $B_2[2]$ -lifting process is counted at least once in  $T_n(A)$  and we have

$$|Q_n(A_{B_2[2]})| \geq |Q_n(A)| - |T_n(A)|.$$

Since  $A_{B_2[2]}$  is a  $B_2[2]$  sequence for any sequence  $A$ , to proof Theorem 1.2 it is enough to prove that there exists a sequence  $A$  such that  $|Q_n(A)| \gg n^\delta$  for some  $\delta > 0$  and for  $n$  large enough and such that  $|T_n(A)| \ll 1$ . We perform these tasks in Propositions 4.1 and 4.2.

**Proposition 4.1.** *With probability 1 we have  $|Q_n(A)| \gg n^{1/11}$  for  $n$  large enough.*

*Proof.* We apply Janson's inequality to  $\Omega = Q_n$  and  $X = |Q_n(A)| = \{\omega \in Q_n : \omega \subset A\}$  where  $A$  is a random sequence in  $\mathcal{S}_m(7/11, S \pmod{N})$ . In Lemma 6.4 we prove that  $\mu_n = \mathbb{E}(Q_n(A)) \gg n^{1/11}$  and in Proposition 6.1 we prove that  $\Delta(Q_n) \ll n^{-2/11}$  for

$$\Delta(Q_n) = \sum_{\substack{\omega, \omega' \in Q_n \\ \omega \sim \omega'}} \mathbb{P}(\omega, \omega' \in A).$$

Thus for  $n$  large enough we have that  $\Delta_n < \mu_n$  and Janson's inequality implies that

$$\mathbb{P}(|Q_n(A)| \leq \mu_n/2) \leq \exp(-\mu_n/12).$$

Then for some  $C > 0$  we can write

$$\sum_n \mathbb{P}(|Q_n(A)| \leq \mu_n/2) < \sum_n \exp(-Cn^{1/11}) < \infty$$

and the Borell-Cantelli lemma implies that with probability 1 we have  $|Q_n(A)| \geq \mu_n/2 \gg n^{1/11}$  for all  $n$  large enough.  $\square$

In the proof of Proposition 4.2 we use several times Lemma 4.1. We first introduce the following families of vectors, whose expected values are bounded in Lemma 6.3.

$$\begin{aligned} (4.1) \quad U_{2r} &= \{\bar{x} = (x_1, x_2) : x_1 + x_2 = r, x_1 \neq x_2\} \\ V_{2r} &= \{\bar{x} = (x_1, x_2) : x_1 - x_2 = r, x_1 \neq x_2\} \\ W_r &= \{\bar{x} = (x_4, x_5, x_6, x_7, x_8) : x_5 + x_6 - x_4 = x_7 + x_8 - x_4 = r, x_i \neq x_j\}. \end{aligned}$$

**Lemma 4.1.** *Let  $X_r$  be any of the three families in (4.1). Then*

$$\mathbb{P}(X_r(A) \text{ contains a 12-d.s.v. for some } r) = o_m(1).$$

*Proof.* Lemma 6.3 implies that  $\mathbb{E}(|X_r(A)|) \ll (r + m)^{-2/11}$  and then apply Proposition 3.1.  $\square$

**Proposition 4.2.** *With probability  $1 - o_m(1)$ ,  $|T_n(A)| \leq 10^{28}$  holds for any  $n$ .*

*Proof.* We claim that

**Claim.** *With probability  $1 - o_m(1)$ ,  $T_n(A)$  does not contain vectorial sunflowers of 12 petals for any  $n$ .*

Assuming the Claim we can apply Corollary 3.1 to the families  $T_n$  to deduce that with probability  $1 - o_m(1)$ , we have that  $|T_n(A)| \leq 8!((8^2 - 8 + 1)12)^8 < 10^{28}$  for all  $n$ . Hence the Claim implies Proposition 4.2.

We prove the Claim for the distinct possible types  $I \subset \{1, \dots, 8\}$  of the vectorial sunflowers in  $T_n(A)$ . The types we analyze below will cover all the cases, as we will explain later.

1.  $I = \emptyset$ . Lemma 6.5  $\implies \mathbb{E}(|T_n(A)|) \ll (n + m)^{-1/11}$ .  
 Proposition 3.1  $\implies \mathbb{P}(T_n(A) \text{ has a 12-d.s.v. for some } n) = o_m(1) \implies$   
 $\implies$  the Claim holds for vectorial sunflowers of type  $I = \emptyset$ .
2.  $|I \cap \{1, 2, 3\}| = 1$ . Suppose that  $I \cap \{1, 2, 3\} = \{1\}$ . The other two cases are similar.  
 If  $T_n(A)$  contains a vectorial sunflower (of 12 petals) of type  $I$  for some  $n$  (denote by  $l_1$  the common first coordinate) then there is a 12-d.s.v.  $\bar{x}_j = (x_{2j}, x_{3j})$ ,  $j = 1, \dots, 12$  such that  $x_{2j} + x_{3j} = n - l_1$ . Thus, for  $r = n - l_1$ ,



$U_{2r}(A)$  contains a 12-d.s.v. and Lemma 4.1 implies the Claim for vectorial sunflowers of this type.

3.  $|I \cap \{1, 4, 5, 6, 7, 8\}| = 1$ . Suppose that  $I \cap \{1, 4, 5, 6, 7, 8\} = \{1\}$ . The other cases are similar.

If  $T_n(A)$  contains a vectorial sunflower (of 12 petals) of type  $I$  for some  $n$  (denote by  $l_1$  the common first coordinate) then there is a 12-d.s.v.  $\bar{x}_j = (x_{4j}, x_{5j}, x_{6j}, x_{7j}, x_{8j})$ ,  $j = 1, \dots, 12$  such that  $x_{5j} + x_{6j} = x_{7j} + x_{8j} = l_1 + x_{4j}$ . Thus, for  $r = l_1$ ,  $W_r(A)$  contains a 12-d.s.v. and Lemma 4.1 implies the Claim for vectorial sunflowers of this type.

4.  $|I \cap \{1, 4, 5, 6\}| = 2$  or  $|I \cap \{1, 4, 7, 8\}| = 2$  or  $|I \cap \{5, 6, 7, 8\}| = 2$ . Suppose that  $|I \cap \{1, 4, 5, 6\}| = 2$ . The other cases are similar. We need to distinguish between two essentially distinct cases:

i)  $I \cap \{1, 4, 5, 6\} = \{1, 4\}$ . If  $T_n(A)$  contains a vectorial sunflower (of 12 petals) of type  $I$  (denote by  $l_1, l_4$  the value of the common coordinates) then there is a 12-d.s.v.  $\bar{x}_j = (x_{5j}, x_{6j})$ ,  $j = 1, \dots, 12$  such that  $x_{5j} + x_{6j} = l_1 + l_4$ . Thus, for  $r = l_1 + l_4$ ,  $U_{2r}(A)$  has a 12-d.s.v. Lemma 4.1 implies the Claim for vectorial sunflowers of this type.

ii)  $I \cap \{1, 4, 5, 6\} = \{1, 5\}$ . If  $T_n(A)$  contains a vectorial sunflower (of 12 petals) of type  $I$  (denote by  $l_1, l_5$  the value of the common coordinates and assume that  $l_1 > l_5$ ) we have that there is an 12-d.s.v.  $\bar{x}_j = (x_{4j}, x_{6j})$ ,  $j = 1, \dots, 12$  such that  $x_{6j} - x_{4j} = l_1 - l_5$ . Thus, for  $r = l_1 - l_5$ ,  $V_{2r}(A)$  contains a 12-disjoint set and Lemma 4.1 implies the Claim for vectorial sunflowers of this type.

The sets  $I$  considered in the previous analysis cover all the possible cases. The point is that if the indexes of one of the equations  $x_1 + x_2 + x_3 = n$ ,  $x_1 + x_4 = x_5 + x_6$ ,  $x_1 + x_4 = x_7 + x_8$ ,  $x_5 + x_6 = x_7 + x_8$  are all in  $I$  except one of them, then a vectorial sunflower of that type  $I$  cannot exist. For example the cases such that  $|I \cap \{1, 2, 3\}| = 2$  are not possible because if two vectors have the same coordinates  $x_1, x_2$ , also  $x_3$  must be the same in both vectors. For example, if  $x_4, x_5, x_6 \in I$  then these coordinates must be the same in all the vectors of the sunflower, but also  $x_1$  should be the same in all of them. The reader can check that the types not studied above are of this kind. Thus we have proved the Claim.  $\square$

5. SIDON BASIS OF ORDER  $3 + \varepsilon$ 

In this section we will prove Theorem 1.3. The proof follows the same steps than the proof of Theorem 1.2 but is a little more involved because we have to distinguish an element  $x_i \leq n^\varepsilon$ .

**5.1. Strategy of the proof of Theorem 1.3.** We start by fixing a cyclic group  $\mathbb{Z}_N$  and a set  $S \subset \mathbb{Z}_N$  satisfying the conditions of Theorem 2.1. Throughout this section we will consider the probabilistic space  $\mathcal{S}_m(\gamma, S \pmod{N})$  with

$$\gamma = \frac{2}{3} + \frac{\varepsilon}{9 + 9\varepsilon}.$$

Indeed we could take any  $\gamma$  with  $\frac{2+3\varepsilon}{3+4\varepsilon} < \gamma < \frac{2+\varepsilon}{3+\varepsilon}$ . We consider the families of sets

$$\begin{aligned} R_n &= \left\{ \omega = \{x_1, x_2, x_3, x_4\} : \text{satisfying the conditions } \text{cond}(R_n) \right\} \quad \text{where} \\ \text{cond}(R_n) &= \begin{cases} x_1 + x_2 + x_3 + x_4 = n, \\ \min(x_1, x_2, x_3, x_4) \leq n^\varepsilon \\ x_i \not\equiv x_j \pmod{N}, \quad 1 \leq i < j \leq 4. \end{cases} \end{aligned}$$

Given a sequence of positive integers  $A$  we define the families:

$$R_n(A) = \left\{ \omega \in R_n : \omega \subset A \right\}.$$

**Definition 7** (Sidon lifting process). *The Sidon lifting process of a sequence  $A$  consists in removing from  $A$  those elements  $a \in A$  such that there exist  $a', a'', a''' \in A$  with  $a + a' = a'' + a'''$ ,  $\{a + a'\} \neq \{a'', a'''\}$ .*

We denote by  $A_{\text{Sidon}}$  the surviving elements of  $A$  after this process.

We define

$$\begin{aligned} B_n(A) &= \left\{ \bar{x} = (x_1, \dots, x_7) : x_i \in A, \bar{x} \text{ satisfies } \text{cond}(B_n) \right\} \quad \text{where} \\ \text{cond}(B_n) &:= \begin{cases} \{x_1, x_2, x_3, x_4\} \in R_n \\ x_1 + x_5 = x_6 + x_7, \quad \{x_1, x_5\} \neq \{x_6, x_7\} \\ x_1 \equiv x_6 \pmod{N}, \quad x_5 \equiv x_7 \pmod{N}. \end{cases} \end{aligned}$$

By a similar argument as the one used in the proof of Theorem 1.2 we can see that  $|B_n(A)|$  is an upper bound for the number of the representations of  $n$  counted in  $R_n(A)$  but destroyed in the Sidon lifting process of  $A$ . Thus,

$$|R_n(A_{\text{Sidon}})| \geq |R_n(A)| - |B_n(A)|.$$

Since  $A_{\text{Sidon}}$  is a Sidon sequence, to prove Theorem 1.3 it is enough to prove that there exists a sequence  $A$  such that  $|R_n(A)| \gg n^\delta$  for some  $\delta > 0$  and for  $n$  large enough, and such that  $|B_n(A)| \ll 1$ .

**Proposition 5.1.** *With probability 1 we have that  $|R_n(A)| \gg n^{\frac{2\varepsilon^2}{9+9\varepsilon}}$  for  $n$  large enough.*

*Proof.* We apply Janson's inequality to  $\Omega = R_n$  and  $X = |R_n(A)| = \{\omega \in R_n : \omega \subset A\}$  where  $A$  is a random sequence in  $\mathcal{S}_m(\gamma, s \bmod N)$ . In Proposition 6.7 we prove that  $\mu_n = \mathbb{E}(R_n(A)) \gg n^{\frac{2\varepsilon^2}{9+9\varepsilon}}$  and in Proposition 6.2 we prove that  $\Delta(R_n) \ll n^{\frac{-3\varepsilon+2\varepsilon^2}{9+9\varepsilon}}$  for

$$\Delta(R_n) = \sum_{\substack{\omega, \omega' \in R_n \\ \omega \sim \omega'}} \mathbb{P}(\omega, \omega' \in A).$$

Thus for  $n$  large enough we have  $\Delta(R_n) < \mu_n$  and Janson's inequality implies that

$$\mathbb{P}(|R_n(A)| \leq \mu_n/2) \leq \exp(-\mu_n/12).$$

Then for some  $C > 0$  we have

$$\sum_n \mathbb{P}(|R_n(A)| \leq \mu_n/2) < \sum_n \exp\left(-Cn^{\frac{2\varepsilon^2}{9+9\varepsilon}}\right) < \infty$$

and the Borell-Cantelli lemma implies that with probability 1 we have  $|R_n(A)| \geq \mu_n/2 \gg n^{\frac{2\varepsilon^2}{9+9\varepsilon}}$  for all  $n$ . This proves Proposition 5.1.  $\square$

In the proof of Proposition 5.2 we use several times Lemma 5.1. We first introduce the following families of vectors, whose expected values are bounded in Lemma 6.6.

$$\begin{aligned} (5.1) \quad U_{2r} &= \{\bar{x} = (x_1, x_2) : x_1 + x_2 = r, x_1 \neq x_2\} \\ U_{3r} &= \{\bar{x} = (x_1, x_2, x_3) : x_1 + x_2 + x_3 = r, x_i \neq x_j\} \\ V_{2r} &= \{\bar{x} = (x_1, x_2) : x_1 - x_2 = r, x_1 \neq x_2\} \\ V_{3r} &= \{\bar{x} = (x_1, x_2, x_3) : x_1 + x_2 - x_3 = r, x_i \neq x_j\}. \end{aligned}$$

**Lemma 5.1.** *Let  $K$  a positive integer such that  $K > 18/\varepsilon^2$ . Then for any of the four families  $X_r$  in (5.1)*

$$\mathbb{P}(X_r(A) \text{ contains a } K\text{-d.s.v. for some } r) = o_m(1).$$

*Proof.* Lemma 6.6 implies  $\mathbb{E}(|X_r(A)|) \ll (r+m)^{\varepsilon/6} \ll (r+m)^{-\varepsilon^2/18}$  and then apply Proposition 3.1.  $\square$

**Proposition 5.2.** *With probability  $1 - o_m(1)$  we have  $|B_n(A)| \ll 1$ .*

*Proof. Claim:* Let  $K$  a positive integer such that  $K > 18/\varepsilon^2$ . Then  $B_n(A)$  does not contain vectorial sunflowers of  $K$  petals for all  $n$ , with probability  $1 - o_m(1)$ .

Assuming the Claim we can apply Corollary 3.1 to the families  $B_n$  to deduce that  $|B_n(A)| \leq 7!((7^2 - 7 + 1)K)^7$  for all  $n$ , with probability  $1 - o_m(1)$ .

We prove the Claim for the distinct possible type  $I \subset \{1, \dots, 7\}$  of the vectorial sunflowers in  $B_n(A)$ . The types we analyze below will cover all the cases. It is clear that if  $I \cap \{1, 2, 3, 4\} = \{1, 2, 3\}$ , then vectorial sunflowers of type  $I$  cannot exist because the conditions on  $B_n$  implies that also the 4th-coordinate is common for all vectors. The same argument works for any  $I$  such that  $|I \cap \{1, 2, 3, 4\}| = 3$  or  $|I \cap \{1, 5, 6, 7\}| = 3$ . Also it is clear that there do not exist vectorial sunflowers of type  $I = [7]$ . Thus we have to consider the types:  $I = \emptyset$ ,  $|I \cap \{1, 2, 3, 4\}| = 1$ ,  $|I \cap \{1, 2, 3, 4\}| = 2$ ,  $|I \cap \{1, 5, 6, 7\}| = 1$ ,  $|I \cap \{1, 5, 6, 7\}| = 2$ .

1.  $I = \emptyset$ . Lemma 6.8  $\implies \mathbb{E}(|B_n(A)|) \ll (n + m)^{-\frac{\varepsilon^2}{18}}$ .

Proposition 3.1  $\implies \mathbb{P}(B_n(A) \text{ has a } K\text{-d.s.v. for some } n) = o_m(1) \implies$   
 $\implies$  the Claim holds for vectorial sunflowers of type  $I = \emptyset$ .

2.  $|I \cap \{1, 2, 3, 4\}| = 1$ . Suppose that  $I \cap \{1, 2, 3, 4\} = \{1\}$ . The other three cases are similar.

If  $B_n(A)$  contains a vectorial sunflower of  $K$  petals of this type for some  $n$  (denote by  $l_1$  to the common first coordinate) we have that there exists an  $K$ -d.s.v.  $\bar{x}_j = (x_{2j}, x_{3j}, x_{4j})$ ,  $\text{Set}(\bar{x}_j) \subset A$ ,  $j = 1, \dots, K$  such that  $x_{2j} + x_{3j} + x_{4j} = n - l_1$ . Thus, for  $r = n - l_1$ ,  $U_{3r}(A)$  contains a  $K$ -d.s.v. and Lemma 5.1 implies the Claim for vectorial sunflowers of this type.

3.  $|I \cap \{1, 2, 3, 4\}| = 2$ . Suppose that  $I \cap \{1, 2, 3, 4\} = \{1, 2\}$ . The other six cases are similar.

If some  $B_n(A)$  contains a vectorial sunflower of  $K$  petals of this type for some  $n$  (denote by  $l_1, l_2$  the common first and second coordinate) we have that there exists an  $K$ -d.s.v.  $\bar{x}_j = (x_{3j}, x_{4j})$ ,  $\text{Set}(\bar{x}_j) \subset A$ ,  $j = 1, \dots, K$  such that  $x_{3j} + x_{4j} = n - l_1 - l_2$ . Thus, for  $r = n - l_1 - l_2$  we have that  $U_{2r}(A)$  contains a  $K$ -d.s.v. and Lemma 5.1 implies the Claim for vectorial sunflowers of this type.

4.  $|I \cap \{1, 5, 6, 7\}| = 1$ . Suppose that  $I \cap \{1, 5, 6, 7\} = \{1\}$ . The other four cases are similar.

If some  $B_n(A)$  contains a vectorial sunflower of  $K$  petals of this type (denote by  $l_1$  the common first coordinate) we have that there exists an  $K$ -d.s.v.  $\bar{x}_j = (x_{5j}, x_{6j}, x_{7j})$ ,  $\text{Set}(\bar{x}_j) \subset A$ ,  $j = 1, \dots, K$  such that  $x_{6j} + x_{7j} - x_{5j} = l_1$ . Thus,

for  $r = l_1$  we have that  $V_{3r}(A)$  contains a  $K$ -d.s.v. and Lemma 5.1 implies the Claim for vectorial sunflowers of this type.

5.  $|I \cap \{1, 5, 6, 7\}| = 2$ . We distinguish two essentially distinct cases:

i)  $I \cap \{1, 5, 6, 7\} = \{1, 5\}$ . The case  $I \cap \{1, 5, 6, 7\} = \{6, 7\}$  is similar.

If some  $B_n(A)$  contains a vectorial sunflower of  $K$  petals of this type (denote by  $l_1, l_5$  the common first and 5 coordinate) we have that there exists an  $K$ -d.s.v.  $\bar{x}_j = (x_{6j}, x_{7j})$ ,  $\text{Set}(\bar{x}_j) \subset A$ ,  $j = 1, \dots, K$  such that  $x_{6j} + x_{7j} = l_1 + l_5$ . Thus, for  $r = l_1 + l_5$  we have that  $U_{2r}(A)$  contains a  $K$ -d.s.v. and Lemma 5.1 implies the Claim for vectorial sunflowers of this type.

ii)  $I \cap \{1, 5, 6, 7\} = \{1, 6\}$ . The case  $I \cap \{1, 5, 6, 7\} = \{5, 7\}$  is similar.

If some  $B_n(A)$  contains a vectorial sunflower of  $K$  petals of this type (denote by  $l_1, l_6$  the common first and 5 coordinate and assume that  $l_1 > l_6$ ) we have that there exists an  $K$ -d.s.v.  $\bar{x}_j = (x_{5j}, x_{7j})$ ,  $\text{Set}(\bar{x}_j) \subset A$ ,  $j = 1, \dots, K$  such that  $x_{7j} - x_{5j} = l_1 - l_6$ . Thus, for  $r = l_1 - l_6$  we have that  $V_{2r}(A)$  contains a  $K$ -d.s.v. and Lemma 5.1 implies the Claim for vectorial sunflowers of this type.

□

## 6. EXPECTED VALUES

We define the quantities:

$$\begin{aligned}\sigma_{\alpha,\beta}(n) &= \sum_{\substack{x,y \geq 1 \\ x+y=n}} x^{-\alpha} y^{-\beta} = \sum_{1 \leq x < n} x^{-\alpha} (n-x)^{-\beta}, \\ \tau_{\alpha,\beta}(n) &= \sum_{\substack{x,y \geq 1 \\ x-y=n}} x^{-\alpha} y^{-\beta} = \sum_{1 \leq x} x^{-\alpha} (n+x)^{-\beta}\end{aligned}$$

and in general

$$\sigma_{\alpha,\beta}(n; m) = \sum_{\substack{x,y > m \\ x+y=n}} x^{-\alpha} y^{-\beta}, \quad \tau_{\alpha,\beta}(n; m) = \sum_{\substack{x,y > m \\ x-y=n}} x^{-\alpha} y^{-\beta}.$$

The next Lemma will be apply later many times. We will write  $\ll^*$  to mean that we are using Lemma 6.1 in an inequality. All  $x_i$  appearing in this section are positive integers.

**Lemma 6.1.** *For any  $\alpha, \beta < 1$  with  $\alpha + \beta > 1$  we have*

$$\begin{aligned}\text{i) } \sigma_{\alpha,\beta}(n; m) &\ll (n+m)^{1-\alpha-\beta}. & \text{iii) } \sigma_{\alpha,\beta}(n) &\ll n^{1-\alpha-\beta}. \\ \text{ii) } \tau_{\alpha,\beta}(n; m) &\ll (n+m)^{1-\alpha-\beta}. & \text{iv) } \tau_{\alpha,\beta}(n) &\ll n^{1-\alpha-\beta}.\end{aligned}$$

*Proof.* If  $n < 2m$ , i) holds because  $\sigma_{\alpha,\beta}(n; m) = 0$ . If  $n \geq 2m$  we have

$$\begin{aligned} \sigma_{\alpha,\beta}(n; m) &\leq \sum_{1 \leq x \leq n/2} x^{-\alpha}(n-x)^{-\beta} + \sum_{n/2 < x < n} x^{-\alpha}(n-x)^{-\beta} \\ &\ll \sum_{1 \leq x \leq n/2} x^{-\alpha}n^{-\beta} + \sum_{n/2 < x < n} n^{-\alpha}(n-x)^{-\beta} \\ &\ll n^{1-\alpha-\beta} \ll (n+m)^{1-\alpha-\beta}. \end{aligned}$$

To prove ii) we distinguish two cases. If  $n < m$  we have

$$\tau_{\alpha,\beta}(n; m) \leq \sum_{x>m} x^{-\alpha}(n+x)^{-\beta} \leq \sum_{x>m} x^{-\alpha-\beta} \ll m^{1-\alpha-\beta} \ll (n+m)^{1-\alpha-\beta}.$$

If  $n \geq m$  we have

$$\begin{aligned} \tau_{\alpha,\beta}(n; m) &\leq \sum_{x>m} x^{-\alpha}(n+x)^{-\beta} = \sum_{m < x < n} x^{-\alpha}(n+x)^{-\beta} + \sum_{x \geq n} x^{-\alpha}(n+x)^{-\beta} \\ &\ll n^{-\beta} \sum_{1 \leq x < n} x^{-\alpha} + \sum_{x \geq n} x^{-\alpha-\beta} \ll n^{1-\alpha-\beta} \ll (n+m)^{1-\alpha-\beta}. \end{aligned}$$

The cases iii) and iv) follow from i) and ii) taking  $m = 0$ . □

**Lemma 6.2.** *Let  $a, b$  be positive integers. Then for any  $\gamma$ ,  $1/2 < \gamma < 1$ ,*

$$\sum_{1 \leq x} x^{-\gamma}(x+a)^{-\gamma}(x+b)^{1-2\gamma} \ll (ab)^{1-2\gamma}.$$

*Proof.* Suppose that  $a < b$  and split the sum:

$$\begin{aligned} S &= \sum_{x \leq b} x^{-\gamma}(x+a)^{-\gamma}(x+b)^{1-2\gamma} + \sum_{x > b} x^{-\gamma}(x+a)^{-\gamma}(x+b)^{1-2\gamma} \\ &\ll b^{1-2\gamma} \sum_{x \leq b} x^{-\gamma}(x+a)^{-\gamma} + \sum_{x > b} x^{1-4\gamma} \ll^* b^{1-2\gamma} a^{1-2\gamma} + b^{2-4\gamma} \ll (ab)^{1-2\gamma}. \end{aligned}$$

□

### 6.1. Expected values in $\mathcal{S}_m(7/11, S \pmod{N})$ .

**Lemma 6.3.** *We have*

- i)  $\mathbb{E}(|U_{2r}(A)|) \ll (r+m)^{-3/11}$ .
- ii)  $\mathbb{E}(|V_{2r}(A)|) \ll (r+m)^{-3/11}$ .
- iii)  $\mathbb{E}(|W_r(A)|) \ll (r+m)^{-2/11}$ .

*Proof.*

$$\begin{aligned}
\mathbb{E}(|U_{2r}(A)|) &= \sum_{\substack{x,y>m \\ x+y=r}} (xy)^{-\gamma} \ll (r+m)^{1-2\gamma} \ll (r+m)^{-3/11}. \\
\mathbb{E}(|V_{2r}(A)|) &= \sum_{\substack{x,y>m \\ x-y=m}} (xy)^{-\gamma} \ll (r+m)^{1-2\gamma} \ll (r+m)^{-3/11}. \\
\mathbb{E}(|W_r(A)|) &\leq \sum_{\substack{x_4,x_5,x_6,x_7,x_8>m \\ x_5+x_6=x_7+x_8=r+x_4}} (x_4x_5x_6x_7x_8)^{-\gamma} \ll \sum_{x_4} x_4^{-\gamma} \left( \sum_{\substack{x,y>m \\ x+y=r+x_4}} (xy)^{-\gamma} \right)^2 \\
&\stackrel{*}{\ll} \sum_{x_4} x_4^{-\gamma} (r+m+x_4)^{2-4\gamma} \stackrel{*}{\ll} (r+m)^{3-5\gamma} \ll (r+m)^{-2/11}.
\end{aligned}$$

□

**Lemma 6.4.**  $\mathbb{E}(|Q_n(A)|) \gg n^{1/11}$  for  $n$  large enough.

*Proof.* We have  $\mathbb{E}(|Q_n(A)|) = \sum_{\{x_1,x_2,x_3\} \in Q_n} \mathbb{P}(x_1, x_2, x_3 \in A) \geq n^{-3\gamma} |Q'_n|$ , where

$$Q'_n = \left\{ \{x_1, x_2, x_3\} \in Q_n : x_i \equiv S \pmod{N}, x_i > m \right\}.$$

We observe that  $S \subset \mathbb{Z}_N$  is such that  $n \equiv s_1 + s_2 + s_3 \pmod{N}$  for some pairwise distinct  $s_1, s_2, s_3$ . We fix  $s_1, s_2, s_3$  and write  $x_i = s_i + Ny_i$  and  $l = \frac{n-s_1-s_2-s_3}{N}$ . Then  $|Q'_n| \geq |Q_n^*|$  where

$$|Q_n^*| = \left| \left\{ \{y_1, y_2, y_3\} : y_1 + y_2 + y_3 = l : y_i > m \right\} \right| \asymp l^2 \gg n^2,$$

if  $l > 10mN$ . Thus,  $\mathbb{E}(|Q_n(A)|) \geq n^{-3\gamma} |Q_n^*| \gg n^{-3\gamma+2} \gg n^{1/11}$  for  $n$  large enough. □

**Proposition 6.1.**  $\Delta(Q_n) \ll n^{-2/11}$ .

*Proof.* If  $\omega \sim \omega'$  with  $\omega, \omega' \in Q_n$ , both sets have exactly one common element, say  $x_1$ . Thus

$$\begin{aligned}
\Delta(Q_n) &= \sum_{\substack{\omega, \omega' \in Q_n \\ \omega \sim \omega'}} \mathbb{P}(\omega, \omega' \subset A) \ll \sum_{\substack{x_1, x_2, x_3, x'_2, x'_3 \\ x_2+x_3=n-x_1 \\ x'_2+x'_3=n-x_1}} (x_1x_2x_3x'_2x'_3)^{-\gamma} \\
&\leq \sum_{x_1 < n} x_1^{-\gamma} \left( \sum_{\substack{x,y \\ x+y=n-x_1}} (xy)^{-\gamma} \right)^2 \stackrel{*}{\ll} \sum_{x_1 < n} x_1^{-\gamma} (n-x_1)^{2-4\gamma} \\
&\stackrel{*}{\ll} n^{3-5\gamma} \ll n^{-2/11}.
\end{aligned}$$

□

**Lemma 6.5.**  $\mathbb{E}(|T_n(A)|) \ll (n+m)^{-1/11}$ .

*Proof.* It is clear that  $\mathbb{E}(|T_n(A)|) = 0$  if  $n < 3m$ , so it is enough to prove that  $\mathbb{E}(|T_n(A)|) \ll n^{-1/11}$ .

We observe that if  $(x_1, \dots, x_8) \in T_n$  then some of these restrictions happens:

- i) All  $x_i$  are pairwise distinct.
- ii)  $x_7 = x_8$  and  $x_1, x_2, x_3, x_4, x_5, x_6, x_7$  are pairwise distinct.
- iii)  $x_4 \in \{x_2, x_3\}$  and  $x_1, x_2, x_3, x_5, x_6, x_7, x_8$  are pairwise distinct.
- iv)  $x_6 \in \{x_2, x_3\}$  and  $x_1, x_2, x_3, x_4, x_5, x_7, x_8$  are pairwise distinct.
- v)  $x_8 \in \{x_2, x_3\}$  and  $x_1, x_2, x_3, x_4, x_5, x_6, x_7$  are pairwise distinct.

In order to simplify these conditions we observe that iv) and v) are essentially the same condition and that  $x_2$  and  $x_3$  play the same role, so iii) can be substituted by  $x_4 = x_2$  and iv) and v) by  $x_6 = x_2$ . Thus we have

$$\begin{aligned} \mathbb{E}(|T_n(A)|) &\ll \sum_{\substack{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8 \\ x_1 + x_2 + x_3 = n \\ x_1 + x_4 = x_5 + x_6 = x_7 + x_8}} (x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8)^{-\gamma} + \sum_{\substack{x_1, x_2, x_3, x_4, x_5, x_6, x_7 \\ x_1 + x_2 + x_3 = n \\ x_1 + x_4 = x_5 + x_6 = 2x_7}} (x_1 x_2 x_3 x_4 x_5 x_6 x_7)^{-\gamma} \\ &+ \sum_{\substack{x_1, x_2, x_3, x_5, x_6, x_7, x_8 \\ x_1 + x_2 + x_3 = n \\ x_1 + x_2 = x_5 + x_6 = x_7 + x_8}} (x_1 x_2 x_3 x_5 x_6 x_7 x_8)^{-\gamma} + \sum_{\substack{x_1, x_2, x_3, x_4, x_5, x_7, x_8 \\ x_1 + x_2 + x_3 = n \\ x_1 + x_4 = x_5 + x_2 = x_7 + x_8}} (x_1 x_2 x_3 x_4 x_5 x_7 x_8)^{-\gamma} \\ &= S_1 + S_2 + S_3 + S_4. \end{aligned}$$

$$\begin{aligned} S_1 &\ll \sum_{x_1, x_4} (x_1 x_4)^{-\gamma} \sum_{\substack{x_2, x_3 \\ x_2 + x_3 = n - x_1}} (x_2 x_3)^{-\gamma} \sum_{\substack{x_5, x_6 \\ x_5 + x_6 = x_1 + x_4}} (x_5 x_6)^{-\gamma} \sum_{\substack{x_7, x_8 \\ x_7 + x_8 = x_1 + x_4}} (x_7 x_8)^{-\gamma} \\ &\stackrel{*}{\ll} \sum_{x_1, x_4} (x_1 x_4)^{-\gamma} (n - x_1)^{1-2\gamma} (x_1 + x_4)^{2-4\gamma} \stackrel{*}{\ll} \sum_{x_1} x_1^{3-6\gamma} (n - x_1)^{1-2\gamma} \stackrel{*}{\ll} n^{5-8\gamma}. \end{aligned}$$

$$\begin{aligned} S_2 &\ll \sum_{x_1, x_4} (x_1 x_4 \frac{x_1 + x_4}{2})^{-\gamma} \sum_{\substack{x_2, x_3 \\ x_2 + x_3 = n - x_1}} (x_2 x_3)^{-\gamma} \sum_{\substack{x_5, x_6 \\ x_5 + x_6 = x_1 + x_4}} (x_5 x_6)^{-\gamma} \\ &\stackrel{*}{\ll} \sum_{x_1, x_4} (x_1 x_4 (x_1 + x_4))^{-\gamma} (n - x_1)^{1-2\gamma} (x_1 + x_4)^{1-2\gamma} \\ &\ll \sum_{x_1} x_1^{-\gamma} (n - x_1)^{1-2\gamma} \sum_{x_4} x_4^{-\gamma} (x_1 + x_4)^{1-3\gamma} \stackrel{*}{\ll} \sum_{x_1} x_1^{2-5\gamma} (n - x_1)^{1-2\gamma} \stackrel{*}{\ll} n^{4-7\gamma}. \end{aligned}$$



$$\begin{aligned}
S_3 &\ll \sum_{x_1, x_2} (x_1 x_2 (n - x_1 - x_2))^{-\gamma} \sum_{\substack{x_5, x_6 > m \\ x_5 + x_6 = x_1 + x_2}} (x_4 x_5)^{-\gamma} \sum_{\substack{x_7, x_8 > m \\ x_7 + x_8 = x_1 + x_2}} (x_7 x_8)^{-\gamma} \\
&\ll \sum_l \sum_{\substack{x_1, x_2 \\ x_1 + x_2 = l}} (x_1 x_2)^{-\gamma} (n - l)^{-\gamma} \sum_{\substack{x_5, x_6 > m \\ x_5 + x_6 = l}} (x_4 x_5)^{-\gamma} \sum_{\substack{x_7, x_8 > m \\ x_7 + x_8 = l}} (x_7 x_8)^{-\gamma} \\
&\stackrel{*}{\ll} \sum_l (n - l)^{-\gamma} l^{3-6\gamma} \stackrel{*}{\ll} n^{4-7\gamma}.
\end{aligned}$$

The estimate of  $S_4$  is more involved. We observe that given  $x_1, x_3, x_4$  the values of  $x_2$  and  $x_5$  are determined by

$$x_2 = n - x_3 - x_1, \quad x_5 = x_4 + 2x_1 + x_3 - n.$$

$$\begin{aligned}
S_4 &\ll \sum_{x_1} \sum_{x_3 < n - x_1} \sum_{x_4} x_1^{-\gamma} (n - x_3 - x_1)^{-\gamma} x_3^{-\gamma} x_4^{-\gamma} (x_4 + 2x_1 + x_3 - n)^{-\gamma} \sum_{\substack{x_7, x_8 \\ x_7 + x_8 = x_1 + x_4}} (x_7 x_8)^{-\gamma} \\
&\stackrel{*}{\ll} \sum_{x_1} \sum_{x_3 < n - x_1} x_3^{-\gamma} x_1^{-\gamma} (n - x_3 - x_1)^{-\gamma} \sum_{x_4} x_4^{-\gamma} (x_4 + 2x_1 + x_3 - n)^{-\gamma} (x_4 + x_1)^{1-2\gamma}.
\end{aligned}$$

Now we apply Lemma 6.2 to the last sum and later we write  $x_3 = n - x_1 - z$ ,  $z > 0$  to get

$$\begin{aligned}
S_4 &\ll \sum_{x_1} \sum_{x_3 < n - x_1} x_3^{-\gamma} x_1^{1-3\gamma} (n - x_3 - x_1)^{-\gamma} (2x_1 + x_3 - n)^{1-2\gamma} \\
&\ll \sum_{x_1} \sum_{z < n - x_1} (n - x_1 - z)^{-\gamma} x_1^{1-3\gamma} z^{-\gamma} (x_1 + z)^{1-2\gamma} \\
&\ll \sum_{x_1} x_1^{2-5\gamma} \sum_{z < n - x_1} (n - x_1 - z)^{-\gamma} z^{-\gamma} \stackrel{*}{\ll} \sum_{x_1} x_1^{2-5\gamma} (n - x_1)^{1-2\gamma} \stackrel{*}{\ll} n^{4-7\gamma}.
\end{aligned}$$

□

## 6.2. Expected values in $\mathcal{S}_m(\frac{2}{3} + \frac{\varepsilon}{9+9\varepsilon}, S \pmod{N})$ .

**Lemma 6.6.** *We have*

$$\begin{aligned}
\text{i) } \mathbb{E}(|U_{2r}(A)|) &\ll (r + m)^{-1/3}, & \text{ii) } \mathbb{E}(|U_{3r}(A)|) &\ll (r + m)^{-\varepsilon/6}. \\
\text{iii) } \mathbb{E}(|V_{2r}(A)|) &\ll (r + m)^{-1/3}, & \text{iv) } \mathbb{E}(|V_{3r}(A)|) &\ll (r + m)^{-\varepsilon/6}.
\end{aligned}$$

*Proof.*

$$\mathbb{E}(|U_{2r}(A)|) = \sum_{\substack{x,y>m \\ x+y=r}} (xy)^{-\gamma} \ll (r+m)^{1-2\gamma} \ll (r+m)^{-1/3}.$$

$$\mathbb{E}(|V_{2r}(A)|) = \sum_{\substack{x,y>m \\ x-y=m}} (xy)^{-\gamma} \ll (r+m)^{1-2\gamma} \ll (r+m)^{-1/3}.$$

$$\begin{aligned} \mathbb{E}(|U_{3r}(A)|) &\leq \sum_{\substack{x,y,z>m \\ x+y+z=r}} (xyz)^{-\gamma} \leq \sum_{z>0} z^{-\gamma} \sum_{\substack{x,y>m \\ x+y=r-z}} (xy)^{-\gamma} \\ &\stackrel{*}{\ll} \sum_z z^{-\gamma} (r-z+m)^{1-2\gamma} \stackrel{*}{\ll} (r+m)^{2-3\gamma} \ll (r+m)^{-\varepsilon/6}. \end{aligned}$$

$$\begin{aligned} \mathbb{E}(|V_{3r}(A)|) &\leq \sum_{\substack{x,y,z>m \\ x+y-z=r}} (xyz)^{-\gamma} = \sum_z z^{-\gamma} \sum_{\substack{x,y>m \\ x+y=r+z}} (xy)^{-\gamma} \\ &\stackrel{*}{\ll} \sum_z z^{-\gamma} (r+z+m)^{1-2\gamma} \stackrel{*}{\ll} (r+m)^{2-3\gamma} \ll (r+m)^{-\varepsilon/6}. \end{aligned}$$

□

**Lemma 6.7.**  $\mathbb{E}(R_n(A)) \gg n^{\frac{2\varepsilon^2}{9+9\varepsilon}}.$

*Proof.* We have  $\mathbb{E}(|R_n(A)|) = \sum_{\{x_1, x_2, x_3, x_4\} \in R_n} \mathbb{P}(x_1, x_2, x_3, x_4 \in A) \geq n^{-(3+\varepsilon)\gamma} |R'_n|$ , where

$$R'_n = \{\{x_1, x_2, x_3, x_4\} \in R_n : x_i \equiv S \pmod{N}, x_i > m\}.$$

We observe that  $S \subset \mathbb{Z}_N$  is such that  $n \equiv s_1 + s_2 + s_3 + s_4 \pmod{N}$  for some pairwise distinct  $s_1, s_2, s_3, s_4$ . We fix  $s_1, s_2, s_3, s_4$  and write  $x_i = s_i + 256y_i$  and  $l = \frac{n-s_1-s_2-s_3-s_4}{N}$ . Then  $|R'_n| \geq |R_n^*|$  where

$$\begin{aligned} |R_n^*| &= \#\{\{y_1, y_2, y_3, y_4\} : y_1 + y_2 + y_3 + y_4 = l, m < \min(y_1, y_2, y_3, y_4) \leq n^\varepsilon/256\} \\ &\asymp n^\varepsilon l^2 \asymp n^{2+\varepsilon} \end{aligned}$$

Thus,  $\mathbb{E}(|R_n(A)|) \geq n^{-(3+\varepsilon)\gamma} |R_n^*| \gg n^{-(3+\varepsilon)\gamma+2+\varepsilon} \gg n^{\frac{2\varepsilon^2}{9+9\varepsilon}}.$

□

**Proposition 6.2.**  $\Delta(R_n) \ll n^{\frac{-3\varepsilon+2\varepsilon^2}{9+9\varepsilon}}.$

*Proof.* We have that

$$\Delta(R_n) = \sum_{\substack{\omega, \omega' \in R_n \\ \omega \sim \omega'}} \mathbb{P}(\omega, \omega' \in A)$$

We split  $\Delta(R_n)$  in several sums according to the common elements of  $\omega$  and  $\omega'$ . We suppose that  $i = 1$  is the index for which  $x_1 \leq n^\varepsilon$ .

1.  $\omega \cap \omega' = x_1$

$$\begin{aligned} \sum_{\substack{x_1, x_2, x_3, x_4, x'_2, x'_3, x'_4 \\ x_1 + x_2 + x_3 + x_4 = n \\ x_1 + x'_2 + x'_3 + x'_4 = n \\ x_1 \leq n^\varepsilon}} (x_1 x_2 x_3 x_4 x'_2 x'_3 x'_4)^{-\gamma} &\ll \sum_{x_1 \leq n^\varepsilon} x_1^{-\gamma} \left( \sum_{\substack{x_2, x_3, x_4 \\ x_2 + x_3 + x_4 = n - x_1}} (x_2 x_3 x_4)^{-\gamma} \right)^2 \\ &\stackrel{*}{\ll} \sum_{x_1 \leq n^\varepsilon} x_1^{-\gamma} (n - x_1)^{4-6\gamma} \ll n^{4-6\gamma+\varepsilon(1-\gamma)}. \end{aligned}$$

2.  $\omega \cap \omega' = x_j$  for some  $j = 2, 3, 4$ . Without loss of generality we consider the case  $x_2 = x'_2$ :

$$\sum_{\substack{x_1, x_2, x_3, x_4, x'_1, x'_3, x'_4 \\ x_1, x'_1 \leq n^\varepsilon \\ x_3 + x_4 = n - x_2 - x_1 \\ x'_3 + x'_4 = n - x_2 - x'_1}} (x_1 x_2 x_3 x_4 x'_1 x'_3 x'_4)^{-\gamma} \ll \sum_{x_2} x_2^{-\gamma} \left( \sum_{x_1 \leq n^\varepsilon} x_1^{-\gamma} \sum_{\substack{x, y \\ x+y = n-x_2-x_1}} (xy)^{-\gamma} \right)^2$$

Now we split the sum in two sums according to  $x_2 \leq n - 2n^\varepsilon$  or  $n - 2n^\varepsilon < x_2 < n$ .

$$\begin{aligned} \sum_{x_2 \leq n-2n^\varepsilon} x_2^{-\gamma} \left( \sum_{x_1 \leq n^\varepsilon} x_1^{-\gamma} (n - x_2 - x_1)^{1-2\gamma} \right)^2 &\ll \sum_{x_2 \leq n} x_2^{-\gamma} \left( \frac{n - x_2}{2} \right)^{2-4\gamma} \left( \sum_{x_1 \leq n^\varepsilon} x_1^{-\gamma} \right)^2 \\ &\stackrel{*}{\ll} n^{3-5\gamma} n^{2(1-\gamma)\varepsilon} \ll n^{(2-2\gamma)\varepsilon+3-5\gamma}. \end{aligned}$$

$$\begin{aligned} \sum_{n-2n^\varepsilon < x_2 \leq n} x_2^{-\gamma} \left( \sum_{x_1 \leq n^\varepsilon} x_1^{-\gamma} (n - x_2 - x_1)^{1-2\gamma} \right)^2 &\stackrel{*}{\ll} \sum_{n-2n^\varepsilon < x_2 \leq n} x_2^{-\gamma} (n - x_2)^{4-6\gamma} \\ &\ll n^{-\gamma} \sum_{n-2n^\varepsilon < x_2 \leq n} (n - x_2)^{4-6\gamma} \ll n^{-\gamma+(5-6\gamma)\varepsilon}. \end{aligned}$$

3.  $\omega \cap \omega' = \{x_1, x_j\}$  for some  $j = 2, 3, 4$ . Without loss of generality we consider the case  $x_1 = x'_1$  and  $x_2 = x'_2$ :

$$\begin{aligned} \sum_{\substack{x_1, x_2, x_3, x_4, x'_3, x'_4 \\ x_1 + x_2 + x_3 + x_4 = n \\ x_1 + x'_2 + x'_3 + x'_4 = n \\ x_1 \leq n^\varepsilon}} (x_1 x_2 x_3 x_4 x'_3 x'_4)^{-\gamma} &\ll \sum_{x_1 \leq n^\varepsilon} x_1^{-\gamma} \sum_{x_2} x_2^{-\gamma} \left( \sum_{\substack{x_3, x_4 \\ x_3 + x_4 = n - x_1 - x_2}} (x_3 x_4)^{-\gamma} \right)^2 \\ &\stackrel{*}{\ll} \sum_{x_1 \leq n^\varepsilon} x_1^{-\gamma} \sum_{x_2} x_2^{-\gamma} (n - x_1 - x_2)^{2-4\gamma} \\ &\stackrel{*}{\ll} \sum_{x_1 \leq n^\varepsilon} x_1^{-\gamma} (n - x_1)^{3-5\gamma} \ll n^{3-5\gamma+\varepsilon(1-\gamma)}. \end{aligned}$$

4.  $\omega \cap \omega' = \{x_j, x_k\}$  for some  $2 \leq j < k \leq 4$ . Without loss of generality we consider the case  $x_2 = x'_2$  and  $x_3 = x'_3$ :

$$\begin{aligned}
& \sum_{\substack{x_1, x_2, x_3, x_4, x'_1, x'_4 \\ x_1+x_2+x_3+x_4=n \\ x'_1+x_2+x_3+x'_4=n \\ x_1, x'_1 \leq n^\varepsilon}} (x_1 x_2 x_3 x_4 x'_1 x'_4)^{-\gamma} \ll \sum_{\substack{x_2, x_3 \\ x_2+x_3 < n}} (x_2 x_3)^{-\gamma} \left( \sum_{x_1 \leq n^\varepsilon} x_1^{-\gamma} (n - x_2 - x_3 - x_1)^{-\gamma} \right)^2 \\
& \stackrel{*}{\ll} \sum_{\substack{x_2, x_3 \\ x_2+x_3 < n-2n^\varepsilon}} (x_2 x_3)^{-\gamma} \left( n^{-\varepsilon\gamma} \sum_{x_1 \leq n^\varepsilon} x_1^{-\gamma} \right)^2 + \sum_{\substack{x_2, x_3 \\ n-2n^\varepsilon \leq x_2+x_3 < n}} (x_2 x_3)^{-\gamma} ((n - x_2 - x_3)^{1-2\gamma})^2 \\
& \ll \sum_{\substack{x_2, x_3 \\ x_2+x_3 < n-2n^\varepsilon}} (x_2 x_3)^{-\gamma} (n^{-\varepsilon\gamma} n^{\varepsilon(1-\gamma)})^2 + \sum_{n-2n^\varepsilon \leq l < n} \sum_{\substack{x_2, x_3 \\ x_2+x_3=l}} (x_2 x_3)^{-\gamma} (n-l)^{2-4\gamma} \\
& \stackrel{*}{\ll} \sum_{\substack{x_2, x_3 \\ x_2+x_3 < n-2n^\varepsilon}} (x_2 x_3)^{-\gamma} n^{\varepsilon(2-4\gamma)} + n^{1-2\gamma} \sum_{n-2n^\varepsilon \leq l < n} (n-l)^{2-4\gamma} \stackrel{*}{\ll} n^{1-2\gamma+\varepsilon(2-4\gamma)} + n^{1-2\gamma+\varepsilon(3-4\gamma)}.
\end{aligned}$$

Observe that if  $\omega \neq \omega'$  it is not possible that they have three common coordinates. Putting  $\gamma = \frac{2}{3} + \frac{\varepsilon}{9+9\varepsilon}$  in each estimate we have that

$$\begin{aligned}
\Delta(R_n) & \ll n^{4-6\gamma+\varepsilon(1-\gamma)} + n^{3-5\gamma+\varepsilon(1-\gamma)} + n^{-\gamma+\varepsilon(5-6\gamma)} + n^{1-2\gamma+\varepsilon(2-4\gamma)} + n^{1-2\gamma+\varepsilon(3-4\gamma)} \\
& \ll n^{\frac{-3\varepsilon+2\varepsilon^2}{9+9\varepsilon}} + n^{\frac{-3-5\varepsilon+2\varepsilon^2}{9+9\varepsilon}} + n^{\frac{-6+2\varepsilon+3\varepsilon^2}{9+9\varepsilon}} + n^{\frac{-3-11\varepsilon-10\varepsilon^2}{9+9\varepsilon}} + n^{\frac{-3-2\varepsilon-\varepsilon^2}{9+9\varepsilon}} \ll n^{\frac{-3\varepsilon+2\varepsilon^2}{9+9\varepsilon}}.
\end{aligned}$$

□

**Lemma 6.8.**  $\mathbb{E}(B_n(A)) \ll (n+m)^{-\frac{\varepsilon^2}{18}}$ .

*Proof.* It is clear that  $\mathbb{E}(|B_n(A)|) = 0$  if  $n < 4m$ , so it is enough to prove that  $\mathbb{E}(B_n(A)) \ll n^{-\frac{\varepsilon^2}{9+9\varepsilon}}$ .

We observe that if  $(x_1, \dots, x_7) \in B_n(A)$  then some of these restrictions holds:

- i) All  $x_i$  are pairwise distinct.
- ii)  $x_6 = x_7$  and all  $x_1, x_2, x_3, x_4, x_5, x_6$  are pairwise distinct.
- iii)  $x_5 \in \{x_2, x_3, x_4\}$  and all  $x_1, x_2, x_3, x_4, x_6, x_7$  are pairwise distinct.
- iv)  $x_6 \in \{x_2, x_3, x_4\}$  and all  $x_1, x_2, x_3, x_4, x_5, x_7$  are pairwise distinct.
- v)  $x_7 \in \{x_2, x_3, x_4\}$  and all  $x_1, x_2, x_3, x_4, x_5, x_6$  are pairwise distinct.

Thus we have

$$\mathbb{E}(|B_n(A)|) \leq \sum' \mathbb{P}(x_1, \dots, x_7 \in A)$$

$(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$   
 $x_1+x_2+x_3+x_4=n$   
 $x_1+x_5=x_6+x_7$   
 $\min(x_1, x_2, x_3, x_4) \leq n^\varepsilon$

and  $\sum'$  means that  $\bar{x}$  satisfies i), ii), iii), iv) or v).

In order to simplify these conditions we observe that iv) and v) are essentially the same condition and that  $x_2, x_3, x_4$  play the same role, so iii) can be substituted by  $x_5 = x_2$  and iv) and v) by  $x_7 = x_2$ . Thus we have that

$$\begin{aligned}
\mathbb{E}(B_n(A)) &\ll \sum_{\substack{x_1, x_2, x_3, x_4, x_5, x_6, x_7 \\ x_1+x_2+x_3+x_4=n \\ x_1+x_5=x_6+x_7 \\ \min(x_1, x_2, x_3, x_4) \leq n^\varepsilon}} (x_1 x_2 x_3 x_4 x_5 x_6 x_7)^{-\gamma} + \sum_{\substack{x_1, x_2, x_3, x_4, x_5, x_6 \\ x_1+x_2+x_3+x_4=n \\ x_1+x_5=2x_6 \\ \min(x_1, x_2, x_3, x_4) \leq n^\varepsilon}} (x_1 x_2 x_3 x_4 x_5 x_6)^{-\gamma} \\
&+ \sum_{\substack{x_1, x_2, x_3, x_4, x_5, x_6 \\ x_1+x_2+x_3+x_4=n \\ x_1+x_2=x_6+x_7 \\ \min(x_1, x_2, x_3, x_4) \leq n^\varepsilon}} (x_1 x_2 x_3 x_4 x_6 x_7)^{-\gamma} + \sum_{\substack{x_1, x_2, x_3, x_4, x_5, x_6 \\ x_1+x_2+x_3+x_4=n \\ x_1+x_5=x_6+x_2 \\ \min(x_1, x_2, x_3, x_4) \leq n^\varepsilon}} (x_1 x_2 x_3 x_4 x_5 x_6)^{-\gamma} \\
&= S_1 + S_2 + S_3 + S_4.
\end{aligned}$$

We write  $S_1 \leq S'_1 + S''_1$  to distinguish the cases  $x_1 \leq n^\varepsilon$  and  $x_2 \leq n^\varepsilon$ .

$$\begin{aligned}
S'_1 &\ll \sum_{x_1 \leq n^\varepsilon} \sum_{x_2 < n} \sum_{x_5} (x_1 x_2 x_5)^{-\gamma} \sum_{\substack{x_3, x_4 \\ x_3+x_4=n-x_1-x_2}} (x_3 x_4)^{-\gamma} \sum_{\substack{x_6, x_7 \\ x_6+x_7=x_1+x_5}} (x_6 x_7)^{-\gamma} \\
&\stackrel{*}{\ll} \sum_{x_1 \leq n^\varepsilon} \sum_{x_2 < n} \sum_{x_5} (x_1 x_2 x_5)^{-\gamma} (n - x_1 - x_2)^{1-2\gamma} (x_1 + x_5)^{1-2\gamma} \\
&\stackrel{*}{\ll} \sum_{x_1 \leq n^\varepsilon} \sum_{x_2 < n} x_1^{2-4\gamma} x_2^{-\gamma} (n - x_1 - x_2)^{1-2\gamma} \ll \sum_{x_1 \leq n^\varepsilon} x_1^{2-4\gamma} (n - x_1)^{2-3\gamma} \\
&\stackrel{*}{\ll} n^{2-3\gamma+\varepsilon(3-4\gamma)} \ll n^{-\frac{\varepsilon}{3+3\varepsilon}+\varepsilon(\frac{1}{3}-\frac{4\varepsilon}{9+9\varepsilon})} \ll n^{-\frac{\varepsilon^2}{9+9\varepsilon}} \ll n^{-\varepsilon^2/18}.
\end{aligned}$$

$$\begin{aligned}
S''_1 &\ll \sum_{x_2 \leq n^\varepsilon} \sum_{x_1 < n} \sum_{x_5} (x_1 x_2 x_5)^{-\gamma} \sum_{\substack{x_3, x_4 \\ x_3+x_4=n-x_1-x_2}} (x_3 x_4)^{-\gamma} \sum_{\substack{x_6, x_7 \\ x_6+x_7=x_1+x_5}} (x_6 x_7)^{-\gamma} \\
&\stackrel{*}{\ll} \sum_{x_2 \leq n^\varepsilon} \sum_{x_1 < n} \sum_{x_5} (x_1 x_2 x_5)^{-\gamma} (n - x_1 - x_2)^{1-2\gamma} (x_1 + x_5)^{1-2\gamma} \\
&\stackrel{*}{\ll} \sum_{x_2 \leq n^\varepsilon} \sum_{x_1 < n} x_1^{2-4\gamma} x_2^{-\gamma} (n - x_1 - x_2)^{1-2\gamma} \stackrel{*}{\ll} \sum_{x_2 \leq n^\varepsilon} x_2^{-\gamma} (n - x_2)^{4-6\gamma} \\
&\stackrel{*}{\ll} n^{4-6\gamma+\varepsilon(1-\gamma)} \ll n^{-\frac{2\varepsilon}{3+3\varepsilon}+\varepsilon(\frac{1}{3}-\frac{\varepsilon}{9+9\varepsilon})} \ll n^{-\frac{-3\varepsilon+2\varepsilon^2}{9+9\varepsilon}} \ll n^{-\varepsilon^2/18}.
\end{aligned}$$

In the estimates of  $S_2$ ,  $S_3$  and  $S_4$  we remove the annoying condition  $\min(x_1, x_2, x_3, x_4) \leq n^\varepsilon$ .

$$\begin{aligned}
S_2 &\leq \sum_{\substack{x_1, x_2, x_3, x_4, x_5, x_6 \\ x_1+x_2+x_3+x_4=n \\ x_1+x_5=2x_6}} (x_1 x_2 x_3 x_4 x_5 x_6)^{-\gamma} \\
&\leq \sum_{x_1, x_2} (x_1 x_2)^{-\gamma} \left( \sum_{x_5} (x_5 (x_1 + x_5)/2)^{-\gamma} \right) \left( \sum_{\substack{x_3, x_4 \\ x_3+x_4=n-x_1-x_2}} (x_3 x_4)^{-\gamma} \right) \\
&\stackrel{*}{\ll} \sum_{x_1, x_2} (x_1 x_2)^{-\gamma} x_1^{1-2\gamma} (n - x_1 - x_2)^{1-2\gamma} \stackrel{*}{\ll} \sum_{x_1} x_1^{1-3\gamma} (n - x_1)^{2-3\gamma} \ll n^{4-6\gamma}.
\end{aligned}$$

$$\begin{aligned}
S_3 &\leq \sum_{\substack{x_1, x_2, x_3, x_4, x_5, x_6 \\ x_1+x_2+x_3+x_4=n \\ x_1+x_2=x_6+x_7}} (x_1 x_2 x_3 x_4 x_6 x_7)^{-\gamma} \\
&\leq \sum_{x_1, x_2} (x_1 x_2)^{-\gamma} \sum_{\substack{x_6, x_7 \\ x_6+x_7=x_1+x_2}} (x_6 x_7)^{-\gamma} \sum_{\substack{x_3, x_4 \\ x_3+x_4=n-x_1-x_2}} (x_3 x_4)^{-\gamma} \\
&\stackrel{*}{\ll} \sum_{x_1, x_2} (x_1 x_2)^{-\gamma} (x_1 + x_2)^{1-2\gamma} (n - x_1 - x_2)^{1-2\gamma} \\
&\ll \sum_l \sum_{\substack{x_1, x_2 \\ x_1+x_2=l}} (x_1 x_2)^{-\gamma} l^{1-2\gamma} (n - l)^{1-2\gamma} \stackrel{*}{\ll} \sum_l l^{2-4\gamma} (n - l)^{1-2\gamma} \stackrel{*}{\ll} n^{4-6\gamma}.
\end{aligned}$$

$$\begin{aligned}
S_4 &\leq \sum_{\substack{x_1, x_2, x_3, x_4, x_5, x_6 \\ x_1+x_2+x_3+x_4=n \\ x_1+x_5=x_6+x_2}} (x_1 x_2 x_3 x_4 x_6 x_7)^{-\gamma} \\
&\ll \sum_{\substack{x_1, x_2 \\ x_1 < x_2 \\ x_1+x_2 < n}} (x_1 x_2)^{-\gamma} \sum_{\substack{x_5, x_6 \\ x_5-x_6=x_2-x_1}} (x_6 x_5)^{-\gamma} \sum_{\substack{x_3, x_4 \\ x_3+x_4=n-x_1-x_2}} (x_3 x_4)^{-\gamma} \\
&\stackrel{*}{\ll} \sum_{\substack{x_1, x_2 \\ x_1 < x_2 \\ x_1+x_2 < n}} x_1^{-2\gamma} (x_2 - x_1)^{1-2\gamma} (n - x_1 - x_2)^{1-2\gamma} \\
&\ll \sum_{x_1 < n/2} x_1^{-2\gamma} (n - 2x_1)^{3-4\gamma} \ll \sum_{x_1 < n/2} x_1^{-2\gamma} (n/2 - x_1)^{3-4\gamma} \stackrel{*}{\ll} n^{3-4\gamma} n^{1-2\gamma}.
\end{aligned}$$

Thus,  $S_2, S_3, S_4 \ll n^{4-6\gamma} \ll n^{-\frac{2\varepsilon}{3+3\varepsilon}} \ll n^{\frac{-\varepsilon^2}{18}}$ . □

**6.3. Acknowledgements.** I am grateful to Igor Shparlinski for useful comments about Theorem 2.2 and to Joel Spencer and Prasad Tetali for provide me a copy of the paper [24] and some informations about Erdős's conjecture. I am also grateful to Rafa Tesoro for a carefully reading of the paper.

## REFERENCES

- [1] Alon N., Shpilka A. and Umans C., *On Sunflowers and Matrix Multiplication*, Computational Complexity, to appear.
- [2] Alon N. and Spencer J., *The Probabilistic Method, Second edition*, Wiley, New York, (2000).
- [3] Cilleruelo J., *Probabilistic constructions of  $B_2[g]$  sequences*, Acta Mathematica Sinica 26–7 (2010).
- [4] Cilleruelo J., *Sidon basis*, arXiv:1304.5351v1
- [5] Cilleruelo J., Kiss S., Ruzsa I. and Vinuesa, C., *Generalization of a theorem of Erdos and Renyi on Sidon sets*, Random Structures and Algorithms, vol 37, n4 (2010).
- [6] Deshouillers J-M. and Plagne A., *A Sidon basis*, Acta Mathematica Hungarica **123** , 3 (2009), 233–238.
- [7] Erdős P., Einige Bemerkungen zur Arbeit von A. Sthör: Gelöste und ungelöste Fragen über Basen der natürlichen Zahlenreihe, J. Reine Angew. Math 197 (1957).
- [8] Erdős P., *The probability method: Successes and limitations*, Journal of Statistical Planning and Inference, 72 (1998), 207–213.
- [9] Erdős P. and Rado R., *Intersection theorems for systems of sets*, Journal of the London Mathematical Society, Second Series 35 (1960), 85–90
- [10] Erdős P. and Renyi A., *Additive properties of random sequences of positive integers*, Acta. Arith., 6 (1960)
- [11] Erdős P., Sarkozy A. and Sós T., *On additive properties of general sequences*, Discrete Mathematics Volume 136, Issue 13, 31 December 1994, Pages 75–99.
- [12] Erdős P., Sarkozy, A. and Sós, T. *On sum sets of Sidon sets I*, Journal of Number Theory, 47 (1994), 329–347.
- [13] Erdős P. and Tetali P., *Representations of integers as the sum of  $k$ -terms*, Random Structures and Algorithms, 1 (1990), 245–261.
- [14] Erdős P. and Turán P., *On a problem of Sidon in additive number theory and on some related problems*, J. London Math. Soc. 16 (1941), 212–215
- [15] Granville A., Shparlinski I.E. and Zaharescu A., *On the distribution of rational functions along a curve over  $\mathbb{F}_p$  and residue races*. J. Number Theory 112, 216–237 (2005).
- [16] Hasse H., *Zur Theorie der abstrakten elliptischen Funktionenkörper. I, II, III*, Crelle's Journal 1936 (175).
- [17] Halberstan H. and Roth, K. F. *Sequences*, Clarendon Press, Oxford 1966.
- [18] Janson S., *Poisson aproximacion for large desviations*, Randon Structures and Algorithms 1 (1990), 221–230.
- [19] Kiss S., *On generalized Sidon sets which are asymptotic bases*. Preprint.
- [20] Kiss S., *On Sidon sets which are asymptotic basis*, Acta Mathematica Hungarica 128 (2010), 46–58.
- [21] Kiss S., Rozgonyi E. and Sandor C. *On Sidon sets which are asymptotic basis of order four*, arXiv:1304.5749v1
- [22] Kim J.H. and Vu V.H., *Concentration of multivariate polynomials and its applications*, Combinatorica 20 (2000), 417–434.

- [23] Ruzsa I., *Solving a linear equation in a set of integers*, Acta Arith. LXV.3 (1993) 259–282.
- [24] Spencer J. and Tetali P., *Sidon Sets with Small Gaps*, Discrete Probability and Algorithms The IMA Volumes in Mathematics and its Applications Volume 72 (1995), 103–109.

INSTITUTO DE CIENCIAS MATEMÁTICAS (ICMAT) AND DEPARTAMENTO DE MATEMÁTICAS, UNIVERSIDAD AUTÓNOMA DE MADRID, MADRID 28049, ESPAÑA.

*E-mail address:* `franciscojavier.cilleruelo@uam.es`